NEURAL POPULATION CODING OF VISUAL MOTION

A Thesis
Presented to
The Academic Faculty

by

Sean Thomas Kelly

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biomedical Engineering

Georgia Institute of Technology
May 2015

NEURAL POPULATION CODING OF VISUAL MOTION

Approved by:

Professor Garrett Stanley, Advisor
School of Biomedical Engineering
Georgia Institute of Technology

Professor Rob Butera
Department of Electrical and
Computer Engineering
Georgia Institute of Technology

Professor Jose-Manuel Alonso
College of Optometry
State University of New York

Professor Robert Liu
Department of Biomedical Engineering
Emory University

Professor Christopher Rozell
Department of Electrical and
Computer Engineering
Georgia Institute of Technology

Date Approved: December 5, 2014

To my wife,

Stephanie Kelly,

without her patience and love I would never have been able to achieve this.

ACKNOWLEDGEMENTS

Frankly there are too many people to thank for their help, support, and inspiration during my time performing this research.

First I want to thank my advisor, Garrett Stanley, for always believing in my work and for never listening to my first pass at things instead waiting to hear the second analysis.

Second I want to thank everyone in the Stanley lab and in the extended Neurolab in general. From whiteboard work with Daniel Millard to bouncing beta figures off of Clare Gollnick and Clarissa Shephard, everyone in the Stanley Lab has been an invaluable collaborator. I also want to thank those that have been both friends and colleagues: Adam Charles and Jeff Bingham. Having their friendship and support has made rough days easier and good days better to have someone to talk science with.

I must acknowledge the intensity and value of my collaboration with the lab of Jose-Manuel Alonso in the College of Optometry of the State University of New York. Although they were the source of all experimental data, rather than simply trading ideas and data in a disconnected fashion, their lab was always involved in design of stimuli and experiments while I was present during the entirety of all experiments which I designed. This collaboration engaged an extensive back-and-forth resulting in higher quality experimental designs as well as experimental outcomes.

Despite understanding very little of the science I was telling them, my parents have supported me throughout my time at Georgia Tech. Without their love and support I could not have continued my research work, and I am unbelievably fortunate for them to want to read all my work and papers no matter how complicated they may

be.

Finally, I must thank my wife Stephanie Kelly. Her patience and support, more than anything else, made it possible for me to accomplish anything. Without her I would have been alone trying to find someone always willing to listen to anything I had to say about my research and my work. For however little I may have managed to teach her about neuroscience she has returned the favor ten-fold in helping me to be a better more responsible person and scientist.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Motion in the outside world forms one of the primary uses of visual information for many animals. The ability to interpret motion quickly and accurately permits interaction with and response to events in the outside world. While much is known about some aspects of motion perception, there is less agreement about how feature selectivity leading to motion perception is actually formed in the convergent and divergent pathways of the visual system. It is even less clear how these classical understandings of motion processing, often driven by artificial stimuli with little resemblance to the outside world, correspond to responses of neurons when using more natural stimuli. In this thesis, we probe these gaps, first by demonstrating that synchronization within the visual thalamus leads to efficient representations of motion (through tuning properties) in primary visual cortex, exploiting precise timing across populations in a unique manner compared to traditional models. We then create a novel "minimally-natural" stimulus with the appearance of an infinite hallway wallpapered with sinusoidal gratings, to probe how such minimally natural features modulate our predictions of neural responses based upon feature tuning properties. Through encoding and decoding models we find that measuring a restricted tuning parameter space limits our ability to capture all response properties but preserves relevant information for decoding. We finish with an exploration of ethologically relevant natural features, perspective and complex motion, and show that even moderate amounts of each feature within or near the classical V1 receptive field changes the neural response from what classical feature tuning would predict and improves stimulus classification tremendously. Together all of these results indicate that capturing information about motion in the outside world through visual stimuli requires a more

advanced model of feature selectivity that incorporates parameters based on more complex spatial relationships.

CHAPTER I

INTRODUCTION

The importance of vision to our everyday sensory perception can not be overstated; neglecting the eyes themselves and subcortical structures, it has been estimated that 15% of the total cortical area (at least in monkeys) is devoted to visual processing (Hubel, 1988). It is not hard to intuitively understand the power that this devotion to vision gets us: within normal everyday lives we can focus on objects across large distances, we can identify objects within fractions of a second, and we can smoothly follow the motion of even fast-moving objects with ease. Performing such tasks with equal speed and generality using our best computing hardware and software still proves an impossibility - we can only scratch the surface of the capabilities of the visual system.

Despite the importance of vision to our everyday lives, our understanding of the mechanisms of vision and perception is surprisingly rudimentary. While we intuitively understand that the human visual system, as well as that of many mammals, is capable of the difficult dual tasks of object recognition and tracking in light of self-motion we have relatively little understanding of how these tasks are actually accomplished. Perhaps the most interesting observation to be made here is that the visual system works with two seemingly fundamental differences compared to our computational solutions to similar problems in algorithmic computer visual systems. The first is that individual sensor elements have access to only part of the outside world and the convergence of this information occurs iteratively over many neuron layers (from periphery to perception) resulting in increasing levels of feature selectivity, whereas computational models often join pixel-wise information across the scene using learned

models of discrete categorization. Secondly the visual system can only convey information through spiking, in stark contrast to many visual models in computer vision that use more continuous representations. From these observations, we come to two questions that drive visual research; how do neurons generate feature selective responses that ultimately culminate in an object-based perception and how do neurons perform these computations using just spiking data? These questions are complicated by the necessity of the solutions to work irrespective of the input stimulus; models that represent the visual system must capture the array of responses observed while using artificial stimuli as well as natural stimuli. This thesis addresses some of these concerns and provides clarity to the way feature selective responses are driven by neural populations and the way these populations represent complex spatial and motion information within a stimulus. We hypothesize that feature selectivity is created when stimuli modulate the synchronization of spike times within neural populations, and that the way this feature selectivity captures the responses of these populations depends strongly on the content and type of scene presented.

Feature selectivity (that is, neurons which respond selectively to particular features of a stimulus) has been an important focus for vision neuroscience ever since it was first observed. As a specific case of feature selectivity, orientation tuning in the visual cortex has been investigated extensively to determine how it originates from un-tuned input. Despite this keen interest these investigations do not accurately pair the biophysical observations of anatomical connectivity with a mapping of that connectivity to functionally observed parameters of orientation selectivity. With the belief that the existing ideas regarding feature selectivity are likely correct but fail to elucidate the details of the system, this thesis provides an explanation for feature selectivity that integrates our understanding of how populations are connected between visual layers with our understanding of the effects of temporal precision among neural populations. I address this problem by implementing realistically connected models

of visual cortex and capturing the timing between spikes that serve as input to these populations. This model is then used to generate observations of feature selectivity (orientation selectivity) that falls within expected parameters. Furthermore because the this model of feature selectivity is completely agnostic as to the underlying stimulus, looking only at short-term temporal correlations on input population spikes, it is applicable to any sensory system and provides insight to far more than just vision. It also potentially describes many different stages of the cortical levels of processing; information from lower stages generates uniquely selective responses in higher stages of processing simply by changing the synchronization of populations of neurons in the lower stages.

What is ultimately at stake is the way we view feature selectivity as an indicator of neural activity. Philosophically if we could measure the responses of a single neuron to infinite combinations of singular parameters (spatial frequency, temporal frequency, etc.) then we could approach the response of a single neuron as having been driven by those parameter combinations in stimuli. In capturing all of these infinite combinations of scene parameters we effectively map an M dimensional space composed of all possible combinations of N parameters, thus implicitly mapping any nonlinearities in their interactions. This is precisely the perspective that visual neuroscience embraced with the use of artificial stimuli of increasing complexity; as we build up our understanding of simple features we simultaneously begin to inject less-simple features. Unknown, though, is a thorough understanding of how well a such a perspective actually captures the responses of neurons in primary visual cortex (one of the major centers involved in the beginnings of actual perceptual translations in vision) to even slightly natural scenes. This thesis tests these combinatorial parameter representations in a very controlled and systematic manner in order to reveal the visual phenomena that actually challenge our ability to represent neural activity. By doing so, we find that feature selectivity with a limited but descriptive parameter

set is generally a fairly poor predictor of single neuron activity but fares better with populations. From this we can begin to understand the kinds of parameters that are actually required to describe neural activity in ethologically relevant contexts.

All of these pieces are embedded in discussions that relate this innate and implicit feature selectivity at different layers of the visual system with the task of evaluating motion in the outside world. Most of the tasks presented involve motion in the presented scene and at least some explicit attempt to retrieve information about this motion. It is important to clarify that motion itself has many causes; the world could be moving, objects could be moving, the camera itself could be moving, or any combination of all three. While any measures of motion presented in this thesis are treated as absolute changes (regardless of possible source) motion estimation within the visual system is a powerful process as it is the only independent measurement we have to verify that our movements had the intended action (did we head in the direction we desired, did we place an object where we intended to?) and as such the ability to translate visual motion into a body-centered reference frame is incredibly important. An understanding of how perspective and boundaries within motion can be used to generate estimates of outside motion, when paired with internal kinesthetic measurements, offers a tremendous representation of internal human locomotion and planning.

1.1 Organization

To understand the relationship between motion sensitivity and neural populations, I have broken this thesis up into five Chapters. Chapter 2 looks at how we can represent the anatomical origins of motion feature selectivity, demonstrated through orientation tuning which provides directionality information about motion, as the sum total of synchronization among un-tuned populations in lower visual system layers (namely

lateral geniculate nucleus). We conclude by showing that such an organization produces maximally efficient transmission of stimulus information on a single neuron level at population synchronization values that match those expected from existing recordings. These results underpin the necessity of convergent populations to transmit accurate information about motion within the outside world as only through this convergence can we convert these un-tuned sensors into a tuned response. In Chapter 3 we expand our analysis to a novel natural-like scene of a sinusoidally-textured hallway that was created and designed specifically for these experiments. We find that using a description of visual neurons based solely on orientation and spatial frequency tuning, we are unable to capture the responses of individual neurons. Despite this we perform moderately well when decoding stimulus parameters, and perform very well classifying different motions within the hallway stimulus. In Chapter 4 we focus this analysis on two specific features common within natural scenes: perspective and motion boundaries. These features are explored with two new generated stimuli that separate each type of feature. Our analysis reveals strong evidence that spatial frequency changes within visual space as well as boundaries between unique motion fields cause some of the observed failures. Models that account for this would be expected to convey significantly more stimulus information. Finally in Chapter 5 we conclude by discussing the implications of these findings for vision research, as well as future directions necessary to strengthen and expand upon these results. Across all our findings, we emphasize that the convergent action of many neurons within a greater population is necessary to permit transmission of motion information through the visual system and that when such populations act in tandem they reliably encode features of stimuli, permitting accurate decoding and determination of stimulus motion. This stimulus motion, through comparison to internal kinematic models, allows us to interpret what visual motion is caused by own navigation and what is caused by independently moving objects.

CHAPTER II

BACKGROUND

2.1 Vision and the Outside World

Our visual interactions with the outside world are guided by a single question: how are we moving in the world and how is the world moving around us? This question provides answers to seemingly trivial tasks such as walking forward (and making sure the world is moving as we expect it to) as well complex tasks such as determining whether an oncoming car is actually going to hit us. A century of electrophysiological exploration of the visual system, investigating numerous areas of the brain and periphery, has converged on the same answer; generally neurons respond only strongly when something changes (Kandel et al., 2000). Lights transition to darks and darks transition to lights, under natural conditions this generally only happens when something in the outside world has moved. Like the pixels on a CCD, though, visual information originates in the brain as simple samples of a time-varying signal with a very limited view on the outside world (Kandel et al., 2000). From these limited samples how does the visual system generate complex information about motion? The central hypothesis of this thesis is that from these singular small samples of the outside world, densely covering the entirety of space, convergent and divergent populations form and cooperatively generate unique representations of aspects of motion like direction and speed.

2.2 Visual System Anatomy

Light enters the eye through the lens, and falls upon the retina at the back of the eye. Photoreceptors transduce the photons into bioelectrical signals, which are propagated

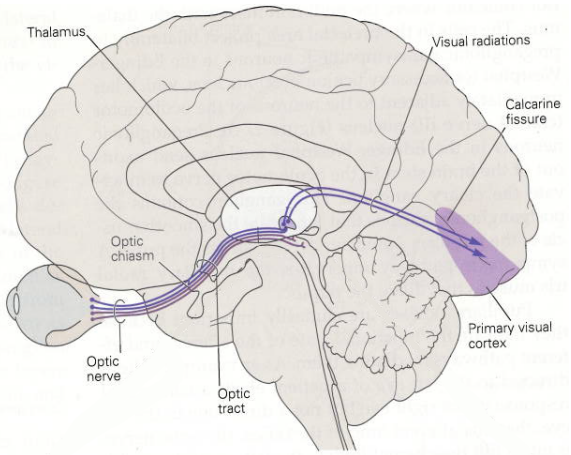

Figure 2.1. The Mammalian Visual System. From the retina, neural information passes through the optic nerve primarily to the lateral geniculate nucleus of the thalamus. From here information then projects to the primary visual cortex. Figure copied from Kandel et al. (2000).

through the layers of neurons in the retina, and projected to the visual thalamus (or lateral geniculate nucleus, LGN). Geniculate relay neurons in turn project primarily to layer 4 of the primary visual cortex (V1). Output layers of primary visual cortex project to higher-order areas of visual cortex and other cortical areas that contain responses to highly processed visual information and guide perception. The early visual system, which we are mostly concerned with here, is shown in Figure 2.1.

Throughout all parts of the early visual system the principle of a topographical map is held in continuity: neurons which are near each other in tissue also represent areas of the outside world that are also near each other, and this relationship is held relatively fixed from retina to visual cortex (Hubel, 1988). Thus any neuron in primary visual cortex, far removed from the neural periphery, will still be surrounded by other neurons receiving visual input from roughly the same spot in visual space. In the retina, retinal ganglion cells (RGCs) gather information from few photoreceptors when near the very center of vision, the fovea, to hundreds of photoreceptors when in the far periphery of vision (Hubel, 1988). In the LGN, neurons receive input from a small number (generally less than three) RGCs and synapse to a number

of neurons in V1 (Hubel, 1988). Classically the LGN has been viewed as simply a relay station, but as this thesis will demonstrate, this view ignores features of LGN transmission in which the population can carry information about stimulus directionality, a key component of motion. The perception of motion in the visual world is incredibly important in our daily lives and interactions with the world, and in mammals visual cortical area MT (medio-temporal, also called V5) is primarily thought to be responsible for this perception both locally and globally. Despite this knowledge of the responsibilities and properties of different areas within the visual system, even for the simplest of visual stimuli neurons within each of these areas have a wide variety of response patterns. Within each of these areas neurons collectively form populations that transmit information about the outside world. From one visual stage to the next, our understanding of exactly how these diverse populations accomplish this transmission is critically limited.

One important concept central to the visual pathway is that of divergence and convergence of connectivity (Hubel, 1988, Kandel et al., 2000). At the retina, a tremendous number of visual receptors converge onto a much smaller number of retinal ganglion cells, on the order of 125 receptors to 1 retinal ganglion cell (depending on proximity to the fovea) (Hubel, 1988). As this information passes through the thalamus the number of connections is essentially preserved but as axons from LGN pass into the visual cortex there is significant divergence and convergence. Multiple LGN neurons will project to a single neuron in the primary visual cortex, yet single LGN neurons will project to multiple neurons within V1 (Purves et al., 2004). This paired convergence and divergence allows for information integration across multiple visual neurons and for a large degree of combinatorial permutations of groups of visual neurons. A practical consequence of this is that synaptic timing becomes exceptionally important (Kandel et al., 2000). Different combinations of visual neurons will present information about the outside world based upon how the spike times within

these combinations align temporally to activate downstream neurons, yet the use of this timing dependence on determining actual functional properties of neurons is still not well understood. This forms the basic functionality of information encoding with populations, convergence and divergence of neural populations between different layers of the visual system encodes a tremendous amount of unique information.

So far I have presented the visual system as a monolithic bundle of neurons that linearly transmit information from one end of the system to another, but this is a misleading view of visual transmission. Starting in the retina, visual information is separated into two distinct pathways based on different layers within the LGN itself and the retinal cells from which their inputs originate. These pathways are called the M (magnocellular) and P (parvocellular) pathways (Kandel et al., 2000, Purves et al., 2004). Both of these pathways converge onto the primary visual cortex and from there the information splits into two incompletely separated paths, the ventral and dorsal pathways. The ventral pathway conveys information to the inferior temporal cortex and follows strongly from the parvocellular pathway to pass information about spatial resolution to permit perception of objects and colors (Kandel et al., 2000, Purves et al., 2004). The magnocellular pathway primarily connects to the dorsal pathway to pass high temporal frequency information, such as object motion, and connects to the posterior parietal cortex.

2.3 Visual Motion

Visual motion, as a broad term, encompasses the movement of the world around us. Whether objects are moving independently or we are actively manipulating them, their temporal change in visual location creates an observable (and with fairly low thresholds) perceivable movement. The overall tuning for these temporal changes are so inherent in the system that studies have found visual signaling (and thus perception of an image) is eventually extinguished if all motion, including endogenous eye shifts,

are removed (Barlow, 1963, Coppola and Purves, 1996). For these reasons, motion can be classified into roughly three categories: motion of the outside world, self motion (caused by active navigation within the outside world), and eye movements (such as saccadic motion caused by repeated rapid and small changes in eye position).

2.3.1 Projections of the Outside World

At the most basic, motion is the movement of individual luminance elements from one visual location to another over some specified period of time (Gibson, 1979). This of course is a very superficial description, but because the real world cannot be broken down into individual elements in the same way computer imagery can (there is, for example, no single distinct pixel element to the real world no matter how small we might define it) talking about motion in natural contexts relies on segregation of the world into objects. Thus a more natural description of motion is the movement of any individual object from one visual location to another over time. While a practical example might be a ball thrown from one hand to the other, experimental examples are more limited and often concern randomly moving small dots or sinusoidal images with a time-varying phase so that it appears to drift in a particular direction. Despite underlying differences in the global image, however, motion itself has a singular effect on visual input and that is that excitation at one retinotopic location will move to a different retinotopic location within some time period. For fast movements this could either be a large distance covered in a long period, or a short distance covered in a short period. Perceiving motion, regardless of which stage of the visual system under consideration, requires some mechanism to bridge these two retinotopic locations such that the excitation is attributed to a single object moving between locations instead of two unrelated objects. More commonly motion occurs along edges either between two differently moving objects or between a moving object and a background. As will be shown, much of our ability to discriminate and estimate motion, even in complex

environments, depends on our ability to predict the direction and speed with which edges are moving within the outside world.

2.3.2 Optic Flow

Describing motion can be challenging, depending on the complexity of the scene and the number of movement sources. The first intuitive understanding came from Gibson (1979) who used a representation of two-dimensional motion called flow perspective (now more frequently referred to as optic flow) to describe the global motion within a visual scene. In the simplest case, he illustrated how translational motion and rotational motions both create coordinated motion velocities that offer implicit information about the underlying motion. While this is a valuable tool for analyzing implicit movement (for example driving a car wherein the movement is global but we are not actively generating the translation) as Gibson points out this coordinated flow field also offers a large amount of information relative to self motion. Convergent flow fields, those wherein motion velocities all point towards some convergence point, indicate translation away from the scene while divergent flow fields indicate translation towards a scene (Gibson, 1979). Head rotations induce a similar velocity field that points in a direction opposite the rotation. Optic flow is also a useful indicator of localized motion and can provide powerful clues as to object segmentation. Just as discontinuities in luminance indicate the presence of edges of objects, discontinuities in the flow field indicate the presence of boundaries between two differently-moving objects.

2.3.3 First Stages of Motion Perception

As will be detailed later, a common perspective of visual motion is that it is primarily perceived in cortical components such as MT and MST - areas which have been distinctly observed to demonstrate responses to motion and **only** motion. This

underestimates the complexity within visual transmission. Varying studies have observed motion-sensitive response in single neurons within the retina (Werblin, 2011, Schwartz et al., 2007) while others have computed accurate representations of motion from populations of neurons within the retina (Frechette et al., 2005, Baccus et al., 2008, Chichilnisky and Kalmar, 2003), both of these indicating that motion information is present and easily accessible within the responses of neurons even at the far periphery. While there is little evidence that these retinal estimates of motion are directly perceptually relevant, anatomical and functional analyses have found that they are incredibly important for coordinating the control of eye location and gaze (Kandel et al., 2000).

2.3.4 Endogenous Eye Movements

Eye movements represent a rather interesting class of motion information because they are endogenously generated and corrupt "true" visual motion information by actively changing the reference frame of the scene. Saccades are the best-understood category, and were originally found to guide the duration of gaze to be highest in areas of strong visual interest in the scene, presumably to maximize their spatial sensitivity at the area centralis or fovea (Purves et al., 2004). Saccades are generated (in part) from the superior colliculus which receives direct synaptic input from retinal ganglion cells, one of many areas which receive visual information outside of the traditional central visual pathway (Purves et al., 2004). Although the superior colluclus itself is not involved in motion perception it seems natural to expect that motion in and of itself is capable of causing saccades given that the superior colliculus both receives strong retinal input and directs saccades to "areas of interest". The result typifies the unique aspect of eye movements: in reaction to a unique and "interesting" sight in the outside world, we move our eyes, thus creating corrupting information on the original motion of interest. A second type of eye motion, smooth pursuit movements,

is less well understood but involves moving the eyes to affix moving objects within a relatively constant retinotopic location (Purves et al., 2004). As with saccadic motion, this also corrupts the underlying motion information, but rather than being a reflexive movement to distinct areas, this control comes from higher-order cortical structures like MT and MST (Kandel et al., 2000).

2.3.5 Cortical Motion Perception

Cortical areas are the prototypical locations for higher-order functions like perception and this is also true for the analysis and observation of motion information within the visual system. The first distinct visual area where motion-induced responses become readily apparent is in the primary visual cortex. The interesting response properties of neurons in primary visual cortex were first isolated by Hubel and Wiesel (1962), who found that the physical rotation of a planar bar of light modulated the firing patterns of neurons in the visual cortex. Since the firing rate of the observed neurons depended most strongly on the orientation of the bar within visual space, this phenomenon was termed "orientation tuning". The importance of the discovery of this orientation tuned response cannot be overstated; since first reported, the concept has formed the keystone of nearly all sensory analysis and modeling in the visual cortex, and potentially other sensory modalities. Many orientation-tuned neurons respond roughly equivalently to any orientations 180 degrees from their preferred orientation (i.e. they respond equivalently to an axis), but in a subset of neurons expressing this tuned response, the tuning is so exclusive that the neuron responds only to a singular direction (Hubel, 1988). These neurons are called "directionally selective" and form the most basic unit of motion-sensitive responses, as it effectively represents a single direction of motion.

Functionality of higher-level areas within the visual system was observed approximately a decade later by Semir Zeki who first isolated directionally selective responses

in what would later be called V5/Area MT (medial temporal) (Born and Bradley, 2005). This area turned out to be more generally selective of direction than primary visual cortex, from which it received direct synaptic input. Further research would separate this cortex into two distinct areas, with MT being associated with motion within a relatively large receptive field and MST (medial superior temporal) being associated more with global motion (Kandel et al., 2000). In both cases, the neurons tend to be more explicitly tuned for motion velocity as compared to motion direction found in V1, and this tuning is often measured by their responses to explicit optic flow stimuli (Smith et al., 2006, Duffy and Wurtz, 1991).

A subset of neurons in area MT also appear to have solved a very important problem in visual motion estimation: the aperture problem (Born and Bradley, 2005, Kandel et al., 2000, Movshon et al., 1983). The aperture problem describes the ambiguity of an oriented bar of light moving in a direction not perpendicular to its surface; for example a diagonal bar moving in any one of the motion vectors indicated in Figure 2.2 would have an identical perception of movement. This occurs because the aperture through which we view the moving bar (similar to the limited view of visual space a single neuron in the visual system) restricts the ability to view motion of the end points of the bar which would provide an unambiguous cue to the true direction. A subset of neurons within area MT have been found to signal the true direction of motion of the bar, however, indicating that they integrate information from outside their own receptive fields to represent an orientation independent motion direction (Born and Bradley, 2005). This is the first location in the visual system that appears to be capable of generating such specific responses at a single neuron level and thus why MT is generally thought to be the first visual area truly associated with estimates of object motion.

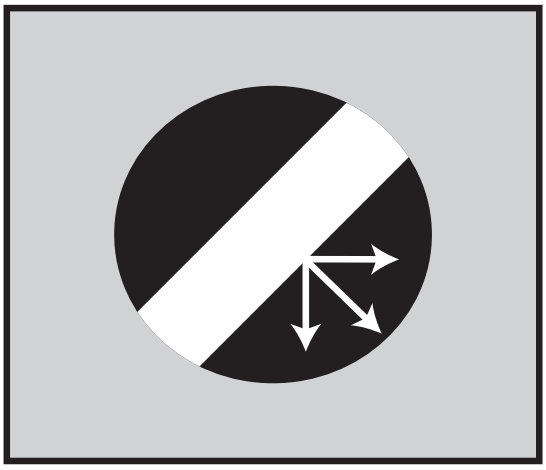

Figure 2.2. The Aperture Problem. When viewing a moving object, here a simple moving bar, viewing the object through an obscuring aperture limits the amount of information we receive about the motion. For this diagonal bar, any of the 3 indicated true motion vectors of the object would have an identical perception of motion which would be perpendicular to the surface of the bar. Thus estimates of motion from single neurons vulnerable to the aperture problem are most strongly influenced by the underlying orientation tuning of the neuron and not the true direction of motion.

2.4 Feature Selectivity and Motion

Although we have already treated the differing kinds of motion perception and sensitivity present within various areas of the visual system, understanding just *how* this sensitivity is generated is important to our knowledge of visual signaling and neural codes. Without discussing specific models of visual information within the visual system, literature has shown that principles underlying motion sensitivity are a more general representation of feature selectivity; we can analyze the way visual neurons are sensitive to more fundamental aspects of motion such as how fast an object is moving and in which direction it is moving. These questions seem rather simplistic, but in general their answers are actually profoundly tied to the "building-up" of a motion representation as visual information moves through the system. We start with a discussion of how the basic representations of visual neurons capture motion feature information and finish with a discussion of more explicit sensitivity to direction and

speed within the cortex.

2.4.1 Visual Receptive Fields

2.4.1.1 Retina and Thalamus

The first line of feature selectivity and the one which expresses the most generalization is the receptive field. A receptive field is the area of visual space from which a given neuron receives excitatory or inhibitory input based on the underlying stimulus luminance (Purves et al., 2004). Retinal ganglion cells and neurons in the LGN have center-surround receptive fields, which means that excitation in the center of the receptive field and dark in the surround is excitatory while the opposite is inhibitory (or vice-versa) (Purves et al., 2004). An example of a center-surround receptive field is shown in Figure 2.3A. When an area of the receptive field is excited by light it is called ON (such as ON-center) and when an area of the receptive field is excited by darkness it is called OFF. Such an arrangement was first identified in retinal ganglion cells by KUFFLER (1953). This arrangement is mostly conserved in the thalamus (LGN) due to the nearly 1:1 convergence of RGC's onto LGN neurons. While a basic representation the visual world, these receptive fields have properties which make them interesting and unique detectors of motion. Both the ON and OFF areas of the center-surround receptive field have temporal dynamics associated with them such that maximal excitation is only achieved when the light source within the area switches "sign"; many ON fields are actually most excited by transitions from dark to light rather than simply the presence of light (Purves et al., 2004). Given that many types of motion are visibly indicated by movements of edges of contrasting illumination, this makes these center-surround RGCs and LGN neurons good indicators of the presence of motion, although they have almost no ability to discriminate any directionality.

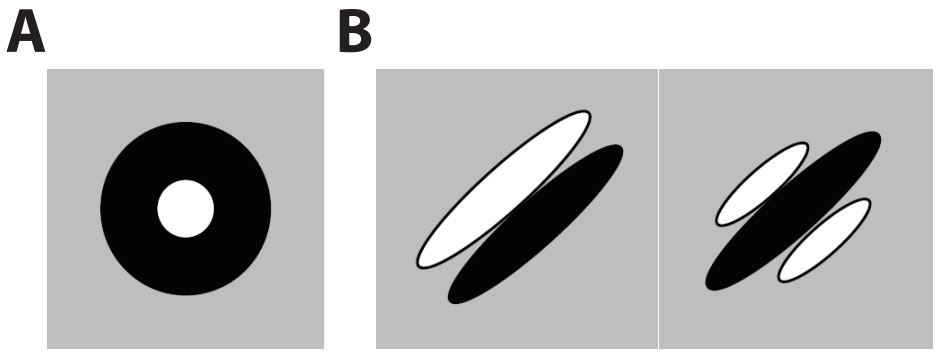

Figure 2.3. Simplified Receptive Field Representations. A. The center surround structures that are common to retinal ganglion cells and LGN neurons with a sample receptive field for an ON-Center configuration. Brightness in the center as well as darkness in the surrounding ring excite the neural response. **B.** Two different styles of receptive field within primary visual cortex. Number of sub-fields as well as elongation in true populations cover a variety of possibilities.

2.4.1.2 Cortical Structures

In the visual cortex the receptive field structure has considerably greater complexity, as can be seen in Figure 2.3B. Not only are the receptive fields generally more elongated, the ON and OFF fields are almost always in a flanking position to each other rather than a complete surround (Hubel, 1988). These two features account for a tremendously increased sensitivity of V1 neurons to particular visual elements that a center-surround receptive field cannot generate. As mentioned previously, cortical neurons are orientation selective and this can be seen in the features of primary visual cortex receptive field., The elongation combined with flanking ON and OFF subfields means that bars of light characterized by strong contrasts (i.e a light bar on a dark background) yield a response that varies strongly with the orientation of the bar-background edge (Hubel, 1988). This preference for particular orientations of light combined with subfields that are (as in the retina and thalamus) most responsive to temporal switches in luminance leads to an expectation of motion selectivity just from the cortical receptive field. Note that the receptive field shown in Figure 2.3B shows a subclass of primary visual cortex neurons know as "simple" cells, so called

for their simplified construction and receptive field shape. Another class of neurons in V1 are " complex" cells which have a receptive field that is much more difficult to visualize as it is constructed from multiple simple cells and thus has a complicated subfield structure (Hubel and Wiesel, 1962). Complex cells form the majority of visual neurons in primary visual cortex (Hubel, 1988).

2.5 Theories Of Motion Selective Responses

Despite the importance of motion estimation in visual scenes in our daily lives, we know comparatively little about how the relevant stimulus information is transmitted and represented through the visual system. As the detection of motion and the detection of motion-based features such as orientation are so intimately related, many of the earliest models of these motion-based features also potentially explain how motion may be encoded in the visual system. One of the earliest and most notable models is the Reichardt Detector (Reichardt, 1987), the first formulation of which actually pre-dated the discovery of directionally selective responses in primary visual cortex. This model, which was built upon response properties within the fly visual system, hypothesizes that visual information from two nearby spatial locations passes through multiple channels: one which is a straight-shot and one which imposes some delay (δ) on the transmission of information as in Figure 2.4A. When the two channels have a very high correlation, that indicates an object in the visual field has moved between the two locations in space with a timing roughly equivalent to δ. This model has a number of failures in that it does not represent actual anatomical connectivity (there is little evidence for delay channels) and also the number of delay lines required for any two pairs of visual location to cover an appropriate range of visual speeds is prohibitively large.

Later models matched anatomical processes and observations much better, particularly in that they invoke filters that are essentially identical to the receptive fields

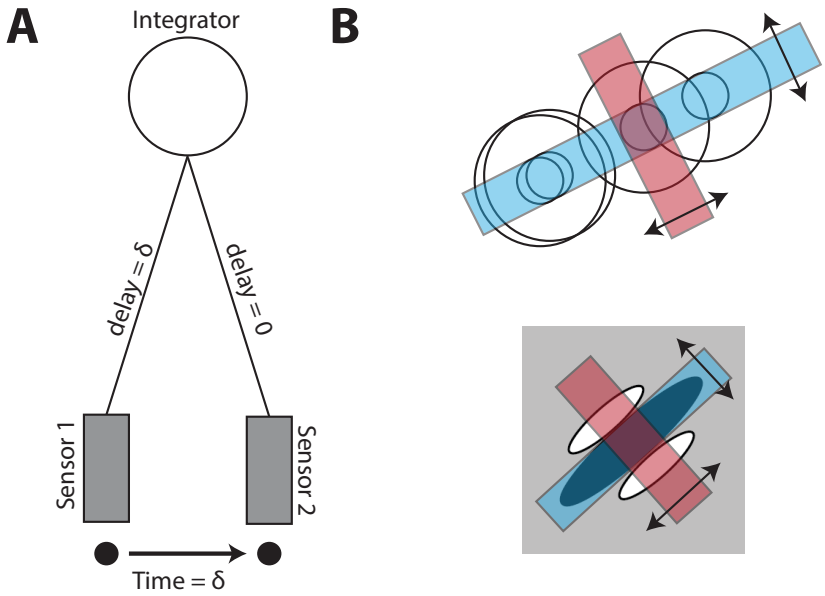

Figure 2.4. Example Motion Detectors. A. A small example Reichardt detector. Two sensors observe a moving object and the time between their observations is δ. If both sensors converge onto an integrating observer, building in a time delay of δ into the transmission of the first sensor ensures that information from both sensors will arrive at the integrator at the same time. Across many different values of δ this construction yields information about direction and speed of motion. **B.** (Top) Collections of LGN-like receptive fields will be simultaneously activated by some directions of motion but not others. (Bottom) Due to their elongated shape single V1 receptive fields will also be strongly activated by only some motion directions.

of the primary visual cortex and thus these models are built on a more explicit usage of orientation and directional selectivity, even across populations, as in Figure 2.4B. For this example (which will appear again later describing Hubel and Weisel's model of orientation tuning) it is clear that across either populations of LGN neurons or elongated receptive fields in primary visual cortex, these elongated structures provide an implicit directionaly selectivity. When bars of light move through them, the rotation (orientation) of that bar will maximally excite these structures at only key orientations.

Evolutions of these models were motion energy models (van Santen and Sperling, 1985, Adelson and Bergen, 1985, Klam et al., 2008) that capture direction and speed of moving objects with a high degree of accuracy. In these models pairs of V1-like filters are combined to an estimate of the direction of movement of part of a stimulus and repeating these pairs of filters at different locations and different sizes across a stimulus yields a rich spatially-varying estimate of local motion direction and speed. Motion energy models, however, describe motion prediction in algorithmic steps that only loosely correspond to existing neural computations. Additional models amend this by explicitly considering the generation of motion-specific information via population calculations (Perrone, 2004, Rust et al., 2006, Nishimoto and Gallant, 2011, Simoncelli and Heeger, 1998) that often pool the directionality information from primary visual cortex to create MT-like motion selectivity. Each of these models shows, in some cases experimentally and in others computationally, that pools of information from V1-like orientation sensors is sufficient to provide estimates of visual motion.

2.6 What do Neural Populations Really Provide?

Noticeably absent from much of the above discussion is the one thing we might expect to be truly important to motion: timing. When visual neurons densely tile the visual space of the outside world, motion in the visual scene creates complex

patterns of activation across local populations of neurons. Rather than trying to "follow" this activation as it moves across this dense tiling (by using constructs similar to Reichardt's delay lines) or skipping lower visual layers and jumping directly to higher-order models built on orientation tuning in primary visual cortex, how might we use this low-level timing information to predict something about stimulus motion? This is the gap we find in the current literature; fairly complex models tell us how retinal ganglion cells as a group capture motion information and what they do with it, and complex models of MT neurons tell us how the cortical population generates motion selectivity, but our **knowledge of how this information is transmitted between these two visual stages is limited.** Throughout this thesis I rely on a single hypothesis: populations of neurons at every level of the visual system collectively represent an important aspect of stimulus motion. It is clear that as we move higher and higher up in these levels, some of the representations are more explicitly clear. However, even for areas in which singular neurons are completely uninformative populations of such neurons are vastly important for effectively capturing stimulus information.

2.7 Experiments

Visual processing has been studied in a variety of model animals including flies, salamanders, birds, cats, primates, and humans. The study of the visual system in cats, however, has a special historical precedence as the primary representations of visual neurons that were discovered by Kuffler and Hubel and Wiesel were both initially found in the cat. These findings have since been validated in primate and human visual systems (Kandel et al., 2000, Purves et al., 2004) although there are small differences between difference species. Complicating these studies is the fact that much of the findings within the visual system (as with many electrophysiology studies of the brain) have been performed under the influence of anesthesia. The effects of

anesthesia on functional observations of neurons and populations are not clear, but it is clear is that without anesthesia many studies would be prohibitively difficult if not impossible. Through anesthesia very precise control over systems is possible; not only does it afford the ability to implant and hold electrodes in consistent locations for long-term recordings with high specificity, it also helps fix the visual system so that retinotopic mappings remain consistent throughout experiments. Without anesthesia principled control over the visual image would be much more difficult and thus observations of visual phenomena becomes harder.

CHAPTER III

POPULATION ENCODING OF FEATURE SELECTIVITY THROUGH SYNCHRONIZATION

3.1 Introduction

It is clear that discrimination of motion information within the visual system is tightly tied to the observation of orientation sensitivity within the visual cortex. All successful models of motion discrimination and MT-like motion selective responses are predicated on the convergence of orientation information within primary visual cortex. One concept missing from this discussion though, is the source of directional selectivity itself. As noted previously, thalamic visual neurons have a center-surround receptive field structure with little to no sensitivity to stimulus motion while primary visual cortex has exquisitely formed receptive field structures that promote strong orientation selectivity. These two areas of visual processing are directly connected and so an immediate and pressing question is, from where does orientation tuning and directional selectivity originate? The fact that the underlying encoding of visual information changes from one area of the visual system to the other implies that this link, in this case the thalamocortical bridge, is a transformative structure for visual information. We have analyzed this previously by observing that synchronization within LGN populations displays strong similarity to orientation tuning in cortex for which it serves as input (Stanley et al., 2012). This timing-based transformation can serve as a general model for how sensory systems convey increasing feature selectivity as the information moves to higher-order brain areas, a model in which the intelligent organization of synaptic inputs is itself an important feature-selective computation.

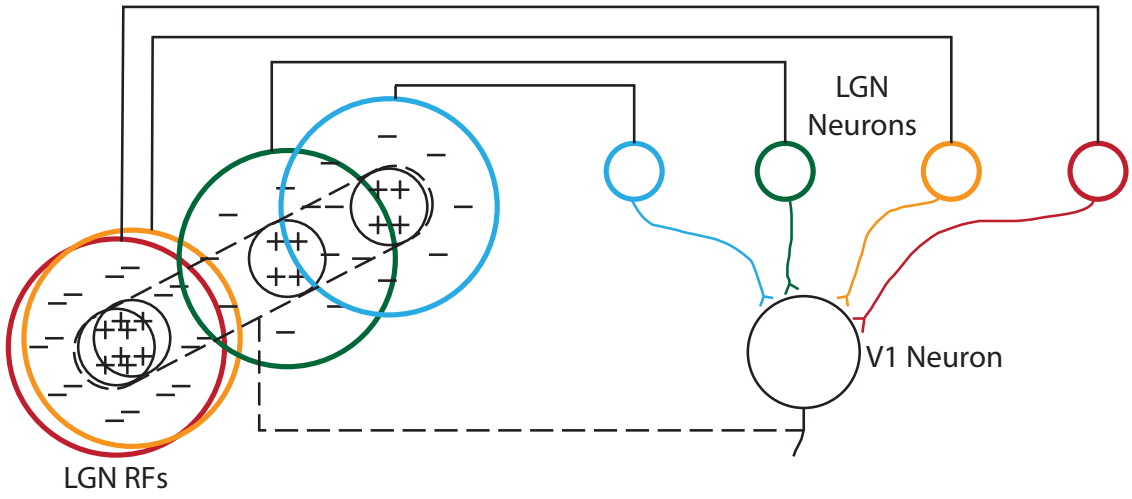

Figure 3.1. Hubel and Wiesel Model of Orientation Selectivity. Hubel and Wiesel proposed a model of V1 orientation selectivity where the selectivity is generated by a strongly oriented population of LGN neurons with convergent synapses onto a single V1 neuron. This convergence generates the strongest excitation when the stimulus is oriented along the length of the population of LGN receptive fields. Adapted from (Hubel and Wiesel, 1962).

3.1.1 Models of Orientation Selectivity

The mechanistic origin of orientation tuning in V1 has been vigorously explored in the literature (Ferster and Miller, 2000, Somers et al., 1995, McLaughlin et al., 2000, Ringach, 2004, Hubel and Wiesel, 1962). In their seminal work, Hubel and Wiesel outlined a conceptual model that involved the projection of LGN neurons along a particular axis of orientation to a common cortical target (Hubel and Wiesel, 1962). Their model is shown in Figure 3.1, which is adapted from their original publication. The model predicts that orientation tuning arises when multiple thalamic neurons with receptive fields arranged along a particular axis all project to the same cortical neuron. This utilizes the idea that simultaneous activation of a series of inputs leads to stronger activation of the downstream neuron, but the underlying width of the proposed thalamic axis is many diameters wide which is unsupported in anatomical observations (Alonso et al., 1996, Reid and Alonso, 1995).

The core connectivity of the Hubel and Wiesel model was subsequently confirmed

in recordings from connected pairs of neurons in LGN and V1 (Tanaka, 1983, Reid and Alonso, 1995, Alonso et al., 2001). Although the relative roles of this feedforward architecture versus cortico-cortico connectivity in sharpening and refining orientation selectivity in such phenomena as contrast-invariance and cross-orientation suppression have been intensely debated (Anderson, 2000, Ferster and Miller, 2000), the thalamic basis for the origin of the basic selectivity is not in dispute, and by its nature implies a role for the timing of thalamic inputs to the cortical target. However, the precise role of timing of thalamic inputs in the downstream cortical orientation selectivity is not known. In the context of the natural visual environment, it has been shown that LGN neurons (individually and across pairs) are temporally precise to a time scale of 10-20 ms, a level that is matched to what is necessary to capture the timescale of changes exhibited in natural scenes (Butts et al., 2007, Desbordes et al., 2008, 2010). Further it has been demonstrated that neurons in the primary visual cortex are extremely sensitive to short intervals between incoming thalamic spikes also on the time scale of approximately 10 ms (Alonso et al., 1996, Usrey et al., 2000, Roy and Alloway, 2001, Azouz and Gray, 2003, Wehr and Zador, 2003, Wilent and Contreras, 2005, Bruno and Sakmann, 2006, Kumbhani et al., 2007, Cardin et al., 2010, Wang et al., 2010b) and that common cortical convergence is most probable when receptive fields overlap (Alonso et al., 1996, Reid and Alonso, 1995). All of these findings collectively suggest that feature selectivity is likely to arise from the modulation of precise timing among overlapping populations of neurons in LGN and that this modulation drives the coactivation of neurons within the populations. Finally, we have recently shown that considering just the coactivation between pairs of electrophysiologically recorded thalamic neurons reveals in many cases extremely sharp orientation tuning even when the receptive fields are highly overlapped (Stanley et al., 2012).

To explore the role of the precise timing of thalamic spiking in the feature selectivity of the downstream cortical neurons to which the thalamus projects, we utilized

experimental population recordings of single units from the LGN region of the visual thalamus in concert with a large-scale thalamocortical model. Specifically, based on anatomical and physiological evidence concerning the convergence of thalamic input to cortical layer 4, we constructed thalamic sub-populations from experimentally recorded thalamic spiking in response to oriented visual stimuli, and systematically controlled the precise timing across the sub-population and its direct impact on the downstream orientation tuning. We found that the conventionally measured tuning sharpness was remarkably invariant over a wide range of peak LGN timing precisions, but the trial-to-trial variability in cortical response was strongly influenced by the timing precision of the LGN inputs. From a decoding perspective of an ideal observer of the cortical response, this complex relationship led to a decreasing error in estimation of orientation with increasing thalamic precision, and a corresponding increase in the information rate, both saturating for peak thalamic precisions of 10-20ms, a finding which was invariant to the overall width of cortical orientation tuning. Taken together, the results here provide a compelling picture for the role of stimulus-driven thalamic synchrony in the emergence of cortical feature selectivity. This feature selectivity is later hypothesized to prove a key role in understanding how motion information is transmitted through the visual system.

3.2 Methods

3.2.1 Ethics Statement

Surgical and experimental procedures were performed in accordance with United States Department of Agriculture guidelines and were approved by the Institutional Animal Care and Use Committee at the State University of New York, State College of Optometry.

3.2.2 Surgical Preparation and Electrophysiological Recordings.

The experimental data collection has been previously described (Stanley et al., 2012). Briefly, single-cell activity was recorded extracellularly in the lateral geniculate nucleus (LGN) of anesthetized and paralyzed male cats, with a total of three animals. As described in (Weng et al., 2005), cats were initially anesthetized with ketamine (10 mg kg^{-1} intramuscular) and acepromazine (0.2mg/kg), followed by propofol (3 mg kg^{1} before recording and 6 mg kg^{-1} h^{-1} during recording; supplemented as needed). A craniotomy and duratomy were performed to introduce recording electrodes into the LGN (anterior, 5.5; lateral, 10.5). Animals were paralyzed with vecuronium bromide (0.3 mg kg^{-1} h^{-1} intravenous) to minimize eye movements, and were artificially ventilated. Using a seven-electrode matrix, layer A geniculate cells were recorded extracellularly. The multielectrode array was inserted into the brain to record from iso-retinotopic lines across the depth of the LGN, using an angle of 25-30 degrees antero-posterior, 2-5 degrees lateral-central. To a multielectrode array (with inter-electrode separation of 254 μm) we attached a glass guide tube with an inner diameter of 300 μm. As the elevation axis is better represented in LGN than the azimuth axis, some of the populations of LGN receptive fields showed greater lateral than vertical scatter in the visual field (Sanderson, 1971). Layer A of LGN was physiologically identified by performing several electrode penetrations to map the retinotopic organization of the LGN and center the multielectrode array at the retinotopic location selected for this study (5-10 degrees eccentricity). While recording, the RASPUTIN software (Plexon, Dallas, TX) was used to capture voltage signals after being amplified and filtered. We isolated single units by independently moving each electrode and the resulting units were spike-sorted online and verified offline using a commercially available algorithm (Plexon, Dallas, TX). Cells were eliminated from this study if they did not have at least 1 Hz mean firing rates in response to all stimulus conditions. Cells were classified as ON or OFF according to the polarity of the receptive

field estimate.

3.2.3 Visual Stimulation.

For each cell, visual stimulation consisted of multiple repetitions of a drifting sinusoidal grating at 0.5 cycles/degree, at either 100% or 64% contrast. The direction of the drifting grating was varied. The orientation of a particular drifting grating was one of eight possible values: 0, 45, 90, 135, 180, 225, 270, 315 degrees. The convention was that a vertically oriented grating drifting rightward was referred to as 0 degrees, a horizontally oriented grating drifting downward was referred to as 90, and so on. The temporal frequency for all datasets was 5 Hz or 4 Hz. The spatial resolution for the drifting gratings was 0.0281 degrees per pixel. All stimuli were presented at a 120 Hz monitor refresh rate.

3.2.4 Generating LGN Population Activity for Model Input.

Biophysiological levels of LGN population synchrony were measured from multiple sets of simultaneous electrophysiological recordings (between 5 and 7 neurons were recorded simultaneously). A cortical neuron is thought to receive approximately 30 LGN inputs (Alonso et al., 2001) but these neurons are substantially more densely arrayed than we can reasonably hope to record with penetrating electrodes. Population response estimates were achieved by expanding the simultaneous recorded neurons into a population of 30 neurons by replicating the recorded responses and then shifting to a new visual location, restricted within the visual space bounded by the original receptive field locations. This restriction resulted in a population that has a receptive field center diameter distribution that is consistent with (Jin et al., 2011) (see Figure 3.2D). To create the population, random shifts were allowed in both the vertical and horizontal directions (i.e. the major and minor axes of the population) but the restrictions placed by the original population layout often required greater

shifts along one or the other axis. For the example in Figure 3.2C, the shift restrictions resulted in a visual space coverage of approximately 5 degrees (horizontal) by 2 degrees (vertical). Shifting responses required knowledge of the timing difference in excitation between the old and the new location, defined as the shift latency. The replicated input spike trains occurred in response to sinusoidal gratings and, due to the regularity in the stimulus, the shift latency was relatively easy to calculate. This shift latency was estimated simply by measuring the timing latency between the maximum excitation at the centroids of the receptive fields at both the original location and the shifted location

$$Latency(\theta) = \frac{\cos(\theta_R)\sqrt{\delta x^2 + \delta y^2} f_s}{f_t} = \frac{\cos(\theta_R) D f_s}{f_t} = \frac{D_\theta f_s}{f_t}$$

where $\sqrt{\delta x^2 + \delta y^2} = D$ is the center to center separation of the original and shifted locations, f_s, f_t represent the spatial (cycle/deg) and temporal (Hz) frequencies (fixed) of the stimulus itself, and θ_R is the angle between the axis connecting the two receptive fields and a line from the shifted location perpendicular to the oriented stimulus bar. A graphical representation of this is in Figure 3.2D. Each newly created neuron is assigned a random trial from all recorded trials of the original neuron and the shift latency value is added to all spike times within that chosen trial. For the representation of this process in Figure 3.2E each neuron received a trial from the appropriate stimulus orientation. As the overarching cortical model, though, expands to a much larger set of orientations than originally recorded from, for consistency each newly created neuron was assigned a trial from the recordings performed with a stimulus at a 0 degree orientation. This allows us to preserve the baseline across-neuron timing changes, while capturing the stimulus-driven timing modulations with our $\sigma_S(\theta)$ parameter, discussed below.

The model was constructed such that all input synapses to the cortical neuron have equal strength and no particular synaptic location (i.e. along the dendrite or at the soma), and accordingly the source of the spikes from within the LGN

population has no effect on the actual model output. Since this is the case, we can estimate the input population auto-correlation by collapsing all LGN spike times into a single vector. The auto-correlation is then calculated by subtracting each spike time from all other spike times and calculating the histogram of these pair-wise interspike intervals. Synchronous populations will have a much higher proportion of small intervals (neglecting stimulus periodicity) than asynchronous populations. The auto-correlations are also appropriately normalized to be between 0 and 1. To smooth the auto-correlation and eliminate correlations caused by the periodicity of the input, a Gaussian was fit to the central 200 ms lags in the correlation. We use timing jitter as a metric of synchrony, which is determined by normalizing the Gaussian fit and locating the lag at which this curve is equal to $1/e$. To relate this number to the PSTH timing jitter (i.e. combined population timing jitter) we must divide by two (see Supplemental in (Butts et al., 2007) for a complete description). In brief, we define a value τ_R which is the "response timescale". This value is equal to the latency at which the auto-correlation equals $1/e$. By construction this has the relationship that $\tau_R = 2\sigma_J$, where σ_J is the timing jitter in the PSTH, our value of interest. This process was performed for all stimulus orientations (in order to maintain phase and timing differences that arise from differences in neuron properties and not just spatial relationships) to describe timing jitter as a function of stimulus orientation. This function was calculated multiple times for different randomly generated populations to estimate the variance that is created by choosing either different visual locations for the component neurons or choosing different recorded trials to represent the neurons in the population.

The observed timing variability in spike times across the population is composed of two aspects; intrinsic neural variability and variability caused by the interaction between the stimulus and the population organization. Our model captured the intrinsic variability by using spike times that were recorded in vivo. On the other

hand, while the grating stimulus always evokes firing in the thalamic neurons the timing differences in spike times from neuron to neuron will vary according to orientation of these gratings and the arrangement of the population itself. We capture this stimulus-evoked timing variability in a parameter $\sigma_S(\theta)$. This parameter, as a function of stimulus orientation, was manually calibrated such that when used with recorded data we could reconstruct the exact plot shown in Figure 3.3C. This procedure allows us to capture both the intrinsic and stimulus-evoked sources of spike timing variability even at orientations for which we were not able to collect data.

3.2.5 Integrate and Fire Model of Direct Synaptic Input to a Cortical Layer 4 Neuron.

All simulations and computations were performed in the Matlab programming language (Mathworks, Inc., Natick, MA) using a 64-node grid computer. The integrate and fire model (Gerstner and Kistler, 2002), illustrated in Figure 3.4A, takes spiking activity from the simulated LGN population as input and outputs cortical membrane potential and the associated cortical spike times. It was assumed that each synapse has equal strength. To create the synaptic input current, an exponentially decaying EPSC of defined amplitude ($A_{EPSC} = 0.05nA$) and time constant ($\tau_{EPSC} = 2ms$) was generated for all spike times in the input LGN population. The EPSCs were summed linearly across all LGN inputs to create a single current input at every simulation time point. The cortical membrane potential was modeled with the following first-order differential equation:

$$\frac{dV_m}{dt} = (R_m I_{EPSC}(t) - (V_m(t) - V_{rest}))(1/\tau_m)$$

where V_m is the membrane potential, I_{EPSC} is the total synaptic current, R_m is the membrane resistance ($100M\Omega$), V_{rest} is the resting potential (-70mV), and τ_m is the membrane time constant (2 ms). The integration was performed using the forward euler method with a step size of 0.05 ms; since the step size is significantly smaller than

any other temporal dynamics or spike timing precision use of a simple euler method is sufficient. When V_m exceeds the threshold membrane potential ($V_{thresh} = -55mV$), a cortical spike is generated by setting the instantaneous potential to 0 mV followed by a 3 ms refractory period at the reset potential of -65 mV. These values are similar to those we have used previously for similar models (Stanley et al., 2012, Wang et al., 2010b). An analysis was performed to determine the approximate sensitivity of the model to each of the above indicated parameters. In general the model is sensitive to parameters that modulate the strength (or efficacy) of input spikes relative to the generated EPSC. Thus the model is sensitive to the EPSC amplitude (A_{EPSC}; effective values 0.05 to 0.1 nA within acceptable ranges) and the EPSC decay (τ_{EPSC}; effective values 2 to 5 ms) while being robust to changes in threshold and reset potentials (V_{reset}, V_{th}). Sensitivity manifests itself as a change between one of three states; impoverished cortical firing, sufficient cortical firing, and strong cortical firing. Impoverished firing results in a peak information per spike (see Figure 3.7C) at very low jitter values (as this maximizes the chance to get any spikes) and strong firing demonstrates no discernible peak information per spike for any particular jitter value (as it results in very wide tuning curves).

3.2.6 Fisher Information.

Taking the perspective of an ideal observer, we approximated the capability of the observer to discriminate between visual stimulus orientations based on cortical activity alone. More specifically, the Fisher information $J(\theta)$ (Zhang and Sejnowski, 1999, Seung and Sompolinsky, 1993, Abbott and Dayan, 1999, Pouget et al., 1999) at each orientation θ captures the discriminability between θ, $\theta + \delta\theta$

$$J(\theta) = E\left[\frac{d}{d\theta}\log(P(r \mid \theta))^2\right]$$

where the expectation is taken with respect to $P(r \mid \theta)$. In the case that the probability is zero, we set $\log(0) = 0$. We calculated the derivative numerically using increments of 1 degree which was the resolution at which the simulations were performed. To reduce the results of this calculation to a single descriptive value, we report the estimator minimum standard deviation, which is related to the Fisher information through the Cramér-Rao lower bound (assuming the estimator is unbiased):

$$\sigma^2_{estimator}(\theta) = \frac{1}{\max(J(\theta))}.$$

As a metric of efficiency with which the cortical output conveys information about the stimulus, we divide the peak output information by the peak spike count with the goal of identifying how much each individual spike contributes to the overall information; higher values indicate each spike is more efficient at conveying information about stimulus features. This established a penalty for higher firing rates, realizing that there is a metabolic cost to generating action potentials.

3.2.7 Estimating Response Distribution.

Response distributions of the cortical firing rate were estimated based on the simulated data, in order to calculate the Fisher Information. The firing rate varied as a function of θ and the distributions are given by $P(r \mid \theta)$. The data were explicitly fit to a Poisson distribution, consistent with previous findings (Tolhurst et al., 1981, Bradley et al., 1987, Vogels et al., 1989, Reich et al., 1997, Baddeley et al., 1997, Gur et al., 1997, Kara et al., 2000, Buracas et al., 1998, Carandini, 2004, Geisler and Albrecht, 1997) as well as explicitly verified for appropriate fitting against our own data:

$$P(r \mid \theta) = \frac{\lambda(\theta)^r \exp(-\lambda(\theta))}{r!}$$

To generate an accurate estimation of the response distributions at a minimum 250 simulation trials were run, with more trials providing no significant change in the estimated distributions. Note that the distributions change as a function of stimulus

orientation, as indicated by $\lambda(\theta)$. Further, in order to create a smooth description of Fisher information it was necessary that the response distributions be smooth functions of θ, as even minor fluctuations in the λ parameter get magnified by differentiation and squaring. To alleviate this, $\lambda(\theta)$ was smoothed with a Gaussian fit which was empirically verified to describe $\lambda(\theta)$ well.

3.3 Results

3.3.1 Spatial Distribution of LGN Populations

Neurons in layer 4 of primary visual cortex are driven by sub-populations of projecting LGN neurons with receptive fields that are highly overlapped, thus representing a relatively limited area of visual space (Jin et al., 2011). Although individual LGN neurons are relatively insensitive to the orientation of drifting sinusoidal gratings, the synchrony across neuron sub-populations is often highly sensitive to the orientation, a product of the relative spatial geometry of the receptive fields and the underlying temporal dynamics of component neurons (Stanley et al., 2012). LGN populations that share a convergent cortical neuron are both large (approximately 30 neurons (Alonso et al., 2001)) and highly overlapped. Since it is not currently possible to record from such dense and numerous clusters in the LGN, we implemented a population-filling method to quantify the synchronization properties of the sub-population. Specifically, in the population-filling method we utilized simultaneous recordings of spiking activity of small sub-populations of LGN neurons whose receptive fields span a small area of visual space (see Methods).

Single unit activity was collected in response to spatiotemporal white noise, and receptive fields (RFs) were mapped using standard spike-triggered averaging (see Methods). The RFs of a pool of simultaneously recorded LGN neurons are shown in Figure 3.2A, where the RF for each neuron is represented as the 20% contour. Note that in this recording, 5 neurons were recorded simultaneously, where each of these

neurons is represented as a different color in the figure. We have previously provided experimental measures of the distribution of receptive field spacing of pairs of LGN neurons monosynaptically connected to a single cortical cell (Alonso et al., 2001) and populations of LGN neurons to a single cortical orientation column (Jin et al., 2011), as shown with the dashed gray curve in Figure 1B. Specifically, this measure provides a probability distribution of the distances between receptive fields, as measured by the distance between the RF centers normalized by the diameter of the larger of the two RFs, referred to here in units of receptive field center diameter (RFCD) - see (Jin et al., 2011).

From experimental data in (Jin et al., 2011), the distribution of separations was modeled as $3.5 * \exp(-2.5x)$, where x is the separation in units of RFCD, which is described only for the range of 0.4 to 2.0. Using the neurons in Figure 3.2A as templates and the relationship in Figure 3.2B (dashed line) as a rule, we filled out the assumed remainder of the population by translating the receptive fields in visual space, creating a dense and accurate convergent LGN population, as shown in Figure 3.2C. The receptive field centers were randomly shifted such that the amount of visual space covered did not change relative to the visual space covered by the original simultaneously recorded population. This method resulted in a distribution of RF separations consistent with previous experimental findings (simulated distribution shown with solid black circles, Figure 3.2B). Note that because the original population was itself elongated in the horizontal axis, the resultant shifts for this population were also mostly horizontal although some receptive field locations also moved vertically. The resultant cluster of receptive fields would be typical for a population that has a major and minor axis as opposed to being more circularly arranged. The resulting aspect ratio of the cluster of RFs in Figure 3.2A is approximately 2.4:1, when measured as the ratio of the longer dimension to the shorter dimension of the area covered by the RF contours. It is important to note that this aspect ratio is lower than the majority

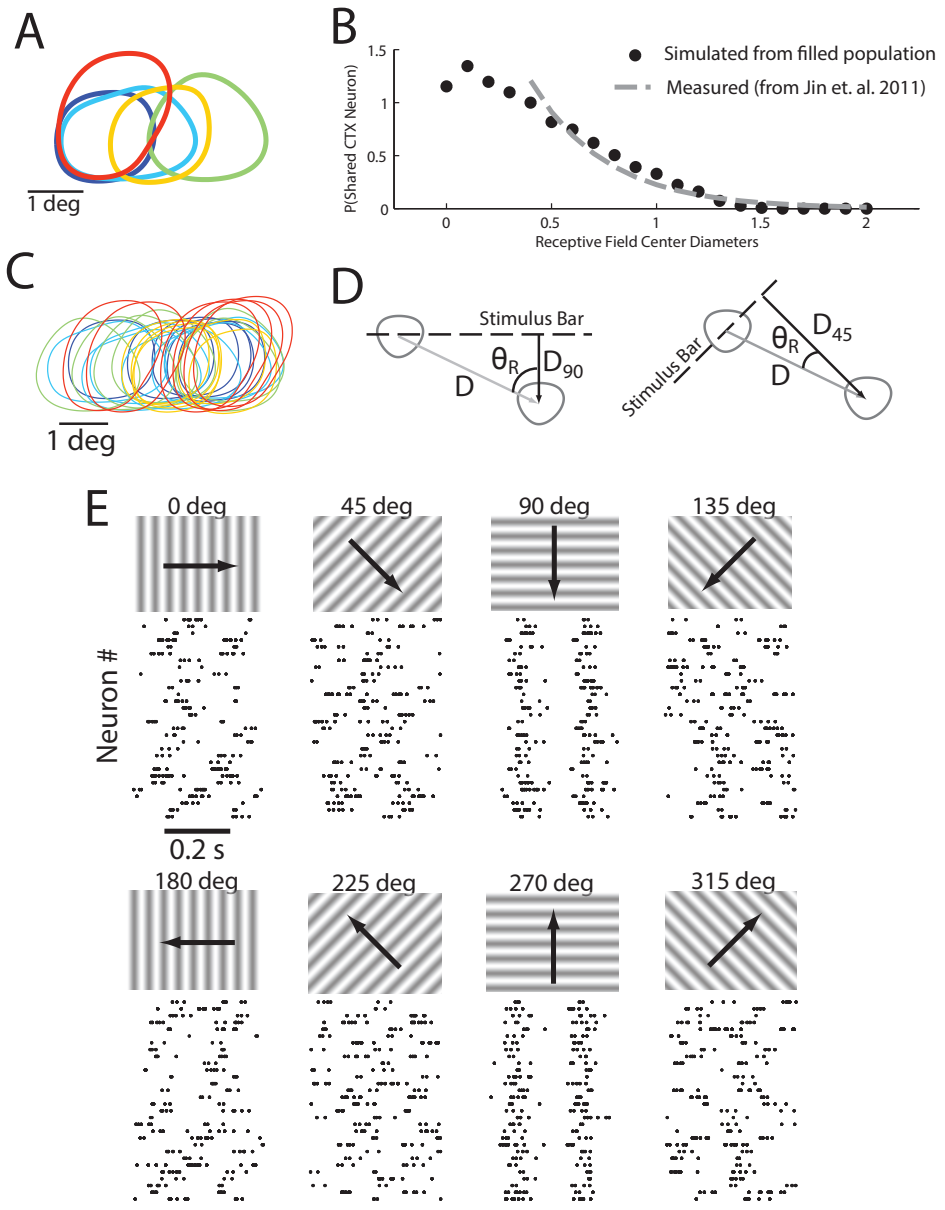

Figure 3.2. Filling in population from recorded neuron receptive fields. **A.** The original simultaneously recorded receptive fields of 5 neurons. **B,C.** The original receptive fields were duplicated and randomly shifted so that the resulting population (**C**) matched the previously measured distribution of RFCD values (**B**) Solid circles indicate RFCD measures from the population in **C**, while the dashed line indicates the expected distribution (see Methods). **D.** The spatial shift in each receptive field describes a particular distance perpendicular to the stimulus orientation that each receptive field shifts; using the spatial and temporal frequencies of the stimulus this can be translated into a timing shift. **E.** Once spike times are appropriately shifted for each neuron in the population, rastergrams reveal spiking alignment only for 90 and 270 degree stimulus orientations.

of existing models (Somers et al., 1995, McLaughlin et al., 2000, Ringach, 2004, Hubel and Wiesel, 1962), where aspect ratios range from 3 to 4 (but see (Ringach, 2004) for a much smaller aspect ratio).

Spiking activity was also collected in response to drifting sinusoidal gratings (0.5 cycles/degree, 5 Hz, 100% contrast - see Methods). The individual LGN neurons had mean firing rates that ranged from 16 to 28 Hz which were relatively insensitive to the stimulus orientation. To generate the population activity in response to the drifting gratings, we utilized the spatially translated RFs as described above, and imposed temporal shifts in the spiking activity based solely on the geometry related to the RF centers, as illustrated in Figure 3.2D. Specifically, a spatial translation of the RF by x degrees horizontally and y degrees vertically imposes a latency shift of the neural response by an amount proportional to the component of the vector connecting the centers of the two RFs orthogonal to the edge of the drifting grating, scaled by the speed of the drift (see Methods). For the collected datasets, spiking activity was collected at each of eight drifting directions with sinusoidal gratings. For each stimulus condition, each randomly placed neuron was assigned a random trial from the original neuron from which it was derived, and the shift latency value was added to all spike times in the chosen trial. In this spirit, we view the trial to trial variability in spiking activity for a single neuron as representative of the across neuron variability on a single trial. The resulting population response at each orientation is shown in Figure 3.2E. For most orientations, spike times within the population uniformly distributed across the entire trial timespan. However, at 90 and 270 degrees, the spike times line up rather precisely between all neurons in the population, reflecting a high degree of synchrony at these orientations.

3.3.2 Physiological Timing Jitter.

The degree of synchrony across this population of neurons is a function of the orientation of the drifting gratings, as well as the variability in spiking timing across neurons within the population. To quantify the synchrony, we used a timing jitter metric, which utilizes the width of the spike-time auto-correlation computed from all spikes in the population (roughly equivalent to the PSTH width). A brief overview of how the auto-correlation is calculated is demonstrated in 3.3A. The collection of spike times across the input population is collapsed into a single spike train, which represents all the projecting thalamic input on the cortical target neuron. This spike train is then used to calculate all of the pair-wise timing differences between every input spike in the population, the histogram of which forms the auto-correlation estimate. There are two values of interest: the population PSTH (with a width of σ_J) and the "response timescale" of the auto-correlation function (given by τ_R). These related values provide us with an approximation for the synchronization within the neural population. When synchrony is high, the spike time auto-correlation has a narrow width and thus there is little jitter. Alternatively, when synchrony is low, the auto-correlation has an increased width and jitter is very high, a property that is demonstrated in Figure 3.3B. From top to bottom in the figure, the level of synchrony in the population increases, spike times become more clustered, and the auto-correlation has a correspondingly decreasing width. Note that each auto-correlation covers the lag range from -400 ms to +400 ms. Each auto-correlation function was fit with a Gaussian between -100 and +100 ms to eliminate any effects of periodicity in response to the drifting sinusoidal grating. The corresponding width of this Gaussian fit was then utilized as the measure of timing jitter. As in (Butts et al., 2007), the timing jitter was defined as the half the latency at which the Gaussian fit is equal to $1/e$ (see Methods and Figure 3.3A). The timing jitter of the population is shown as a function of the stimulus orientation in Figure 3.3C, where the random sampling of

single trials of the template neuron was repeated 50 times. At the most asynchronous stimulus orientations (in this case perpendicular to the elongated axis of the RFs of the population), the timing jitter was approximately 100 ms. At the preferred orientations, when synchrony was maximized, the timing jitter was approximately 24 ms. The timing jitter as a function of stimulus orientation was fit with a Gaussian function (gray dashed line in Figure 3.3C) and exhibited a characteristic tuning width of approximately 31 degrees (standard deviation), a finding which was consistent for two of the three animals. In the third animal there was an insufficient number of strongly-driven neurons with identical polarities (ON- versus OFF-center) to allow for a reasonable reconstruction of a population with more than 2 or 3 neurons. With so few neurons, the population displayed more and more properties of the response of a single neuron as opposed to a rough average of multiple neurons and the overall orientation tuning decreased as the population approached the orientation-agnostic response properties of a single input neuron. To determine the generality of our findings here, we utilized other metrics from previously published studies, with a focus on the reliability method used in (Schreiber et al., 2003) which is easily adaptable to population data. We found that qualitatively the results were similar to our own findings; just as jitter decreases in our sample population at 90 and 270 degrees (Figure 3.3C) the reliability across all the neurons in the population is significantly higher at 90 and 270 degrees. We thus expect that the synchronization observed across all neurons in the population is not affected by the metric chosen to measure it.

By construction, the degree of synchrony across the population of neurons in Figure 3.2D is a function of the orientation of the drifting gratings and across neuron variability in spiking, independent from geometry. The across neuron variability in timing thus set the lower bound of timing jitter in Figure 3.3C. To more fully explore the role of synchrony in shaping the feature selectivity in the downstream cortical response, we effectively replaced the across-neuron variability in spike timing with

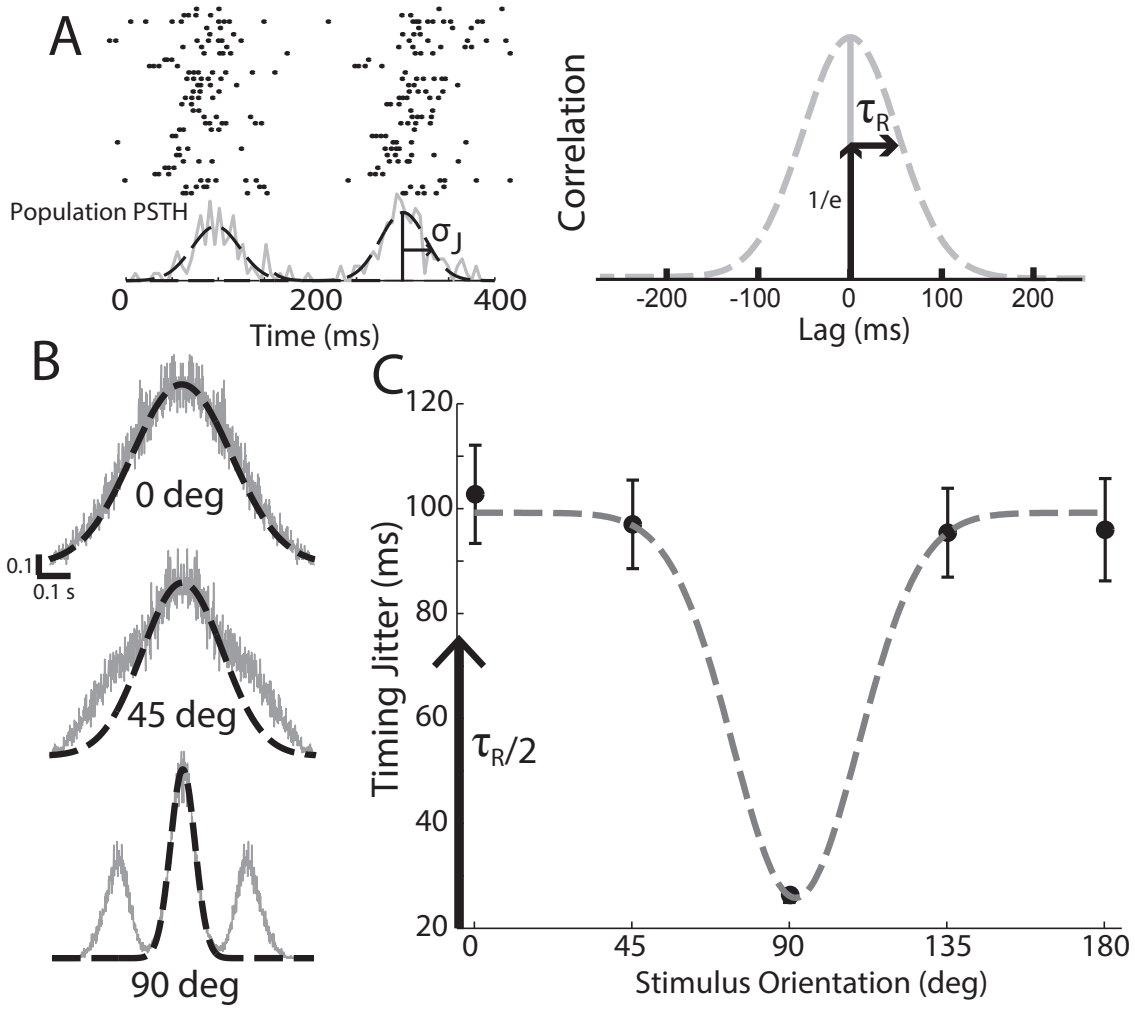

Figure 3.3. Timing jitter is defined by the spike-time auto-correlation width. A. Spike timing auto-correlations come from the spike times across the entire population, collapsed to a single spike train. This can be represented by a PSTH with a particular defined width σ_J which represents timing jitter in ms. The resultant auto-correlation also has a defined width τ_R and this value is the lag at which the auto-correlation is equal to $1/e$, assuming the auto-correlation is appropriately normalized. By construction, $\tau_R = 2\sigma_J$ (see Methods). **B.** Example spike time auto-correlation widths (fit to -0.1 s to 0.1 s with a Gaussian) at non-preferred, moderately preferred, and highly preferred orientations (top to bottom). **C.** The timing jitter is defined through the width of these Gaussian fits and decreases as the stimulus orientation nears the preferred orientation. Black circles indicate measurements taken from recorded data and arranged as in Figure 1, and the dashed gray line indicates a Gaussian fit. Error bars are standard deviation over multiple simulations of the population.

variability under our control. Specifically, we utilized a single trial spike-train for a template neuron and introduced the latency associated with the translation of the receptive field as in Figure 3.2D, but subsequently added variability to each spike time in the form of a Gaussian random variable with zero mean and variance $\sigma_S^2(\theta)$. So long as the population firing rate reaches a particular minimum mean level it does not matter which template neuron is chosen to provide the spike train; we found that nearly all neurons from all three animals provided consistent simulations of cortical activity. Using a single trial has the effect of removing the effects of variable spike count across trials for a particular neuron in addition to providing the exact control over the timing jitter. Of key importance is the value $\sigma_S(\theta)$, which is the stimulus-dependent component of timing jitter (see Methods for expanded description). This value is related to but not equal to the timing value measured from the full populations; $\sigma_S(\theta)$ represents the underlying stimulus-based modulations to synchrony that give rise to the full timing jitter relationship shown in Figure 3.3C. This timing variability quantity $\sigma_S(\theta)$ was parameterized as a Gaussian function of θ and was manually tuned to reproduce the population timing variability curve in Figure 3.3C. From here on out, when we refer to "minimum timing jitter" we are referring to the minimum value of $\sigma_S(\theta)$ that occurs at the preferred orientation.

3.3.3 Cortical Orientation Tuning.

To determine how different levels of input synchrony affect the downstream cortical response and the corresponding feature selectivity, we simulated the cortical layer 4 neuron response to the drifting gratings at different orientations. The previously described populations were used as input to this model, modulating the minimum value of $\sigma_S(\theta)$ to cover a range of 6 to 40 ms of population timing jitter. To model the cortical neuron, we used a biophysically inspired integrate and fire model — illustrated in 3.4A — that generates a continuous membrane potential and corresponding firing

activity, similar to that in (Wang et al., 2010b) and (Stanley et al., 2012) - see Methods. In brief, the model lumps all input spike times together in a common spike train, laying down a superimposed EPSC for each input spike (all of which thus have equal weighting). This model is represented by the differential equation

$$\frac{dV_m}{dt} = (R_m I_{EPSC}(t) - (V_m(t) - V_{rest}))(1/\tau_m)$$

with a fixed parameter set to determine the point by point membrane potential and with a fixed time step of 0.05 ms. Membrane potential traces show a clear stimulus-driven modulation (Finn et al., 2007, Priebe and Ferster, 2005, Anderson et al., 2001) that increases in amplitude towards the population's preferred orientation when averaged over 1000 trials, as shown in Figure 3.4B. Single trial responses, with the exception of the nonphysiological mechanics of the hard reset, match typical recordings from cortical neurons using examples from Carandini & Ferster (Carandini and Ferster, 2000) as a primary source for comparison. Further, the tuning properties (firing rate and tuning half-width at half-height) match reported values, as will be shown later. The reset mechanics did not adversely affect the accuracy of the results as the spiking statistics and tuning curves were consistent with experimental observations. Cortical spike counts, as shown in Figure 3.4C rastergrams, increased dramatically as the stimulus approached the preferred orientation, and the underlying stimulus driven events became very clear. Again, these spike count rastergrams are representative of what would be expected from cortical neurons, although this is easier to see in the cortical tuning curves.

By construction of the thalamic input, the model generated cortical responses that exhibited orientation selectivity. Although the original experimental data was collected only for 8 grating orientations, the parameterized construction described in Figure 3.3 allowed simulation at an arbitrarily fine grain (chosen to be at 1 degree increments here). The resulting mean cortical firing rate across all orientations for a minimum jitter of 6 ms is shown in Figure 3.5A, which is stereotypical of recorded

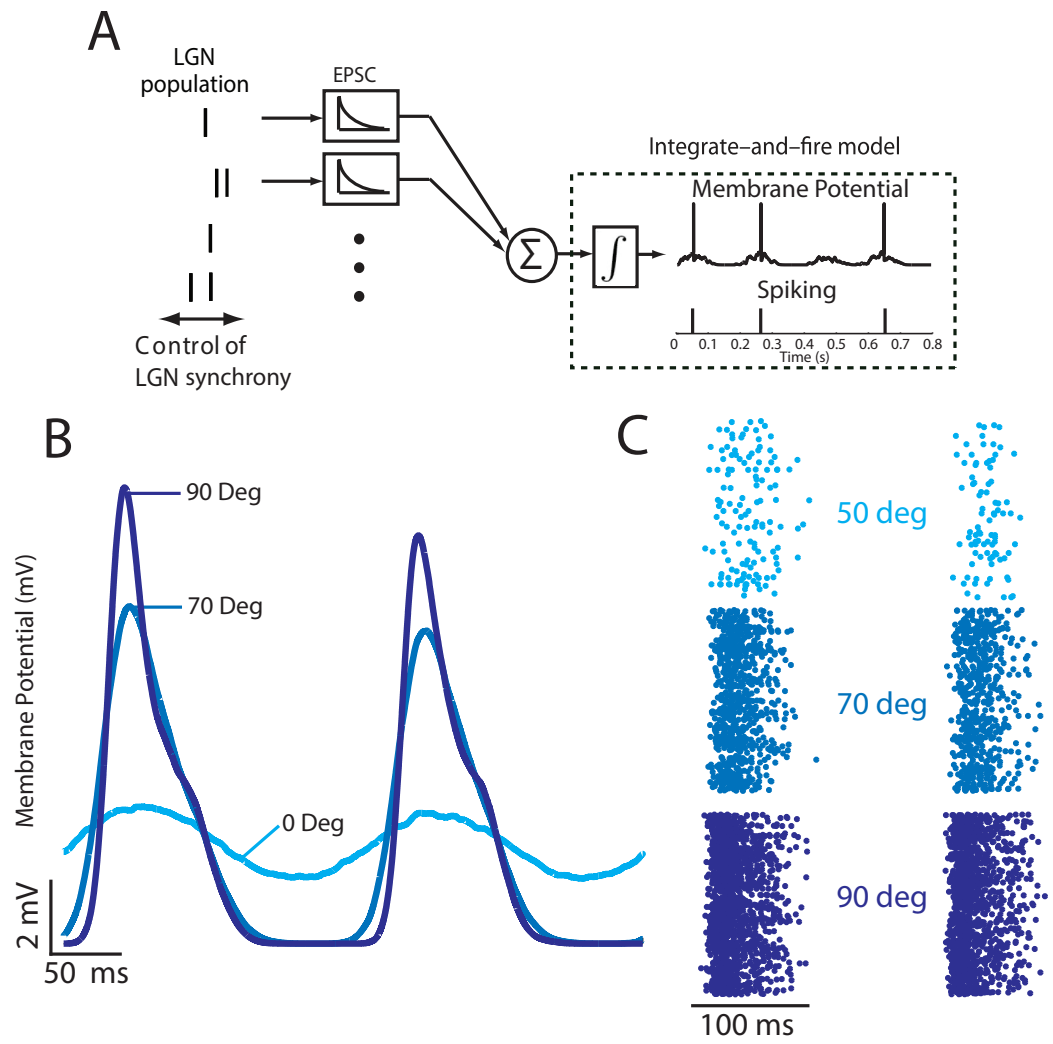

Figure 3.4. Model and simulated output characteristics. A. The model imposes simple control over input spike synchrony and uses a leaky integrate-and-fire construction to determine membrane potential and output spike times. **B.** The simulated cortical membrane potential has an amplitude that is strongly affected by the stimulus orientation, but also a mean value that changes with orientation due to reset characteristics. **C.** Orientations that are closer to the preferred orientation produce dramatically increased numbers of spikes.

responses of neurons in the primary visual cortex (Carandini and Ferster, 2000), with higher firing rates possible when using different neurons for thalamic spike times. The cortical firing rate as a function of stimulus orientation was fit with a local Gaussian over a 180 degree span, as shown with the dashed curve. The parametric fits for each of a range of minimum jitter cases are shown in Figure 3.5B. The colors indicate decreasing levels of synchrony with dark red representing high synchrony (6 ms of jitter) and dark blue representing low synchrony (40 ms of jitter). The overall magnitude of the cortical response decreased with increasing amounts of jitter, as reflected in the overall amplitude of the tuning curves. The sharpness of orientation tuning is quantified through the half-width at half-height (HWHH) of the tuning curve (Carandini and Ferster, 2000, Miller and Troyer, 2002). Consistent with reported values for firing rate, the HWHH tuning width for firing rate was approximately 15 to 16 degrees and was relatively insensitive to the LGN input synchrony (Figure 3.5C) up until 35 ms of input jitter at which point the tuning width increases by approximately 1.5 degrees. These values are on the very low end of expected tuning widths compared to our own observed distributions in Chapter 5 with a mean of 40.3 ± 19.4 degrees but is much closer those from literature (Anderson, 2000, Carandini and Ferster, 2000, Miller and Troyer, 2002). Carandini & Ferster (Carandini and Ferster, 2000) noted that due to experimental limitations they cannot discriminate half-widths less than 17 degrees, a value that they find for almost all recorded neurons. On the other hand different studies (Rose and Blakemore, 1974, Gizzi et al., 1990) have reported tuning widths with significant numbers of neurons with small (10-15 degree) tuning widths. Note that the primary results of the analysis were relatively invariant to the actual tuning width, as we demonstrate in Figure 3.8.

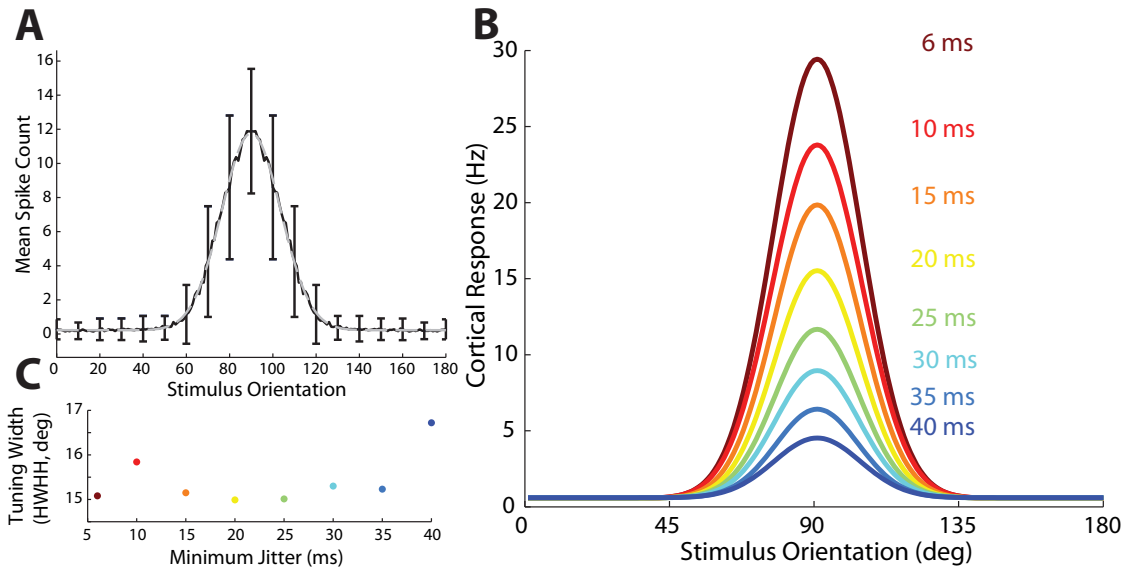

Figure 3.5. Tuned output of cortical model. A. Example tuning curve (black line) at 6 ms of minimum jitter is fit very well by a Gaussian curve (gray dashed line). Standard deviation is illustrated at 10 degree increments, revealing sometimes significant variance in output spike count. In general this reflects the variability of the input spike counts. **B.** The integrate and fire cortical model outputs tuning curves that are well-described by a Gaussian model with an amplitude that decreases with increasing minimum jitter (dark red: 6 ms, dark blue: 40 ms). **C.** The tuning width varies over a small range across the entire range of minimum jitter values simulated.

3.3.4 Statistics of Orientation Tuning.

The tuning curve is illustrative to see how well a particular stimulus orientation drives a cortical neuron but by itself it does not convey any context as to how well the cortical neuron transmits information about the stimulus. Synchrony clearly modulates the overall amplitude of this tuning but it is unclear how it modulates the transmission of the underlying stimulus information. The ability of an ideal observer of neural activity to extract meaningful information regarding the features of a visual stimulus depends not only on the shape of the tuning curve, but also on the variability of the cortical response and how this variability changes with the stimulus feature. The statistics of the cortical response are summarized in Figure 3.6. In Figure 3.6A, the underlying relationship between the mean and variance of the cortical spike count for all stimulus orientations (each individual dot) is illustrated. The relationship clearly demonstrates an increase of spike count variance relative to spike count mean with a slope of approximately 3, which begins to drop when the input is relatively synchronous (6-10 ms of jitter). The variance begins to drop at extreme levels of synchrony as the decreased amount of added timing variance approaches the size of the integration window of the model, and higher synchrony values effectively make the spike count more deterministic. With respect to the relationship between the mean and variance of the cortical response, experimental results have been variable, exhibiting both sub- and supra-linear variability (Tolhurst et al., 1981, Bradley et al., 1987, Vogels et al., 1989, Reich et al., 1997, Baddeley et al., 1997, Gur et al., 1997, Kara et al., 2000, Buracas et al., 1998, Carandini, 2004, Geisler and Albrecht, 1997). So while the orientation tuning width was relatively invariant to the level of synchrony, as shown in Figure 3.5C, the increased level of synchrony was accompanied by an increased mean firing rate, and thus an increased variance, the effects of which are not immediately obvious from the perspective of an ideal observer. Figure 3.6B shows the corresponding spike count distributions for the tuning curves in Figure 3.5B, for

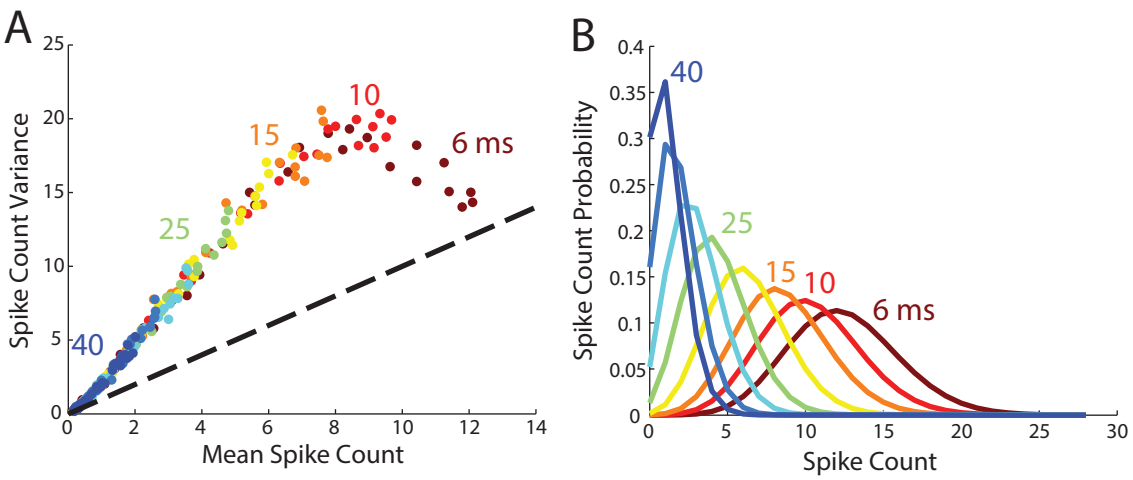

Figure 3.6. Synchrony does not affect the relationship between and mean and variance of output, but does affect discriminability. A. Across all values of synchrony the mean and variance increase in roughly the same linear pattern; each dot is a stimulus orientation from 0 to 180. At high synchrony values relationship is ultimately violated as the spike count variance plateaus, when the timing variance is smaller than the integration window. Jitter values (in units of ms) are indicated next to the dots that represent the simulation results corresponding to those minimum jitter values. **B.** Each curve shows the spike probability distribution at the preferred orientation. Increasing synchrony shifts the spike count distributions away from the origin, giving more freedom to spread and making adjacent orientations more distinguishable (not shown).

the preferred stimulus orientation (90 degrees). The spike count distribution changed dramatically as input synchrony decreased, with asynchronous inputs pinning spike count distributions at the origin and restricting the discriminability at adjacent distributions, a problem not encountered for highly synchronous inputs. From these results we might qualitatively expect that increasing synchrony would lead to increases in information because synchronization appears to give response distributions a greater range over which to vary with stimulus orientation.

Results from both the mean-variance relationship and the per-synchrony peak spike count response distributions thus lead to conflicting expectations on what level of input population synchrony would drive the maximum amount of information about

stimulus orientation. In order to solve this inconsistency we must implement a metric that describes concisely how discriminable different stimulus orientations are and determine the effect input synchrony has on cortical information transfer. Fisher information quantifies the degree to which response distributions are discriminable, and thus, provide unambiguous information about stimulus features captured in the response distributions. The simplest understanding of Fisher information in the context of the problem here is that it represents the derivative of the tuning curve with respect to the stimulus orientation; regardless of the underlying firing statistics, the peak Fisher information will occur near orientations where the derivative of the tuning curve is highest.

3.3.5 Maximum Information Is Modulated By Changes in Input Population Synchrony.

We use the peak amount of information across all stimulus orientations for each level of input synchrony as the metric for the capacity for any particular neuron to inform estimations about the stimulus orientation. By itself the absolute amount of information is an unintuitive quantity. With the goal of determining how synchrony changes the capabilities of cortical neurons to decode specific stimulus features, it is more natural to look at properties of the feature estimator. The inverse of Fisher information is the Cramér-Rao lower bound, a theoretical lower bound on the variance of a maximum-likelihood estimator; decreases in this quantity yield estimates that are more precise and have more confidence. Under the assumption that the stimulus orientation estimator is unbiased, lower estimator variance guarantees lower estimator error. Since we could directly calculate Fisher information in our model, we could also determine what this lower bound was, as shown in Figure 3.7A. The estimator standard deviation decreased nonlinearly with increasing synchrony, covering a range of relatively precise estimates to very imprecise estimates with a notable saturation at around 20 ms of jitter; synchrony higher than this does not yield rapid

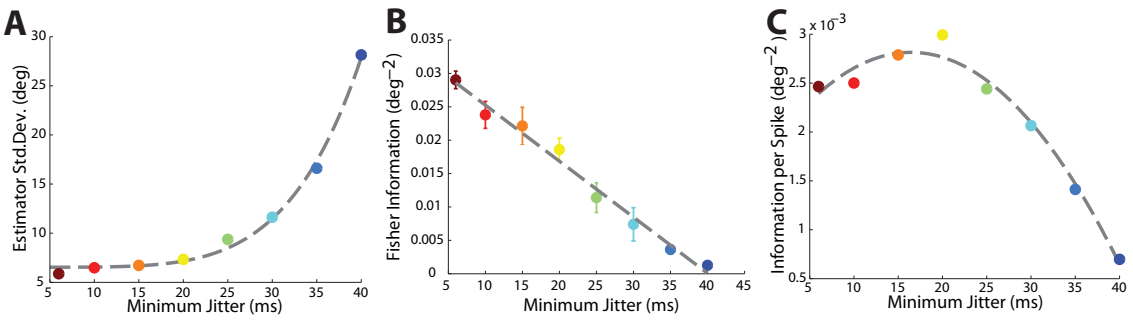

Figure 3.7. Information efficiency peaks as synchrony increases. A. Estimator standard deviation monotonically decreases as the minimum jitter of the input decreases. **B.** The absolute amount of information decreases approximately linearly with increasing minimum jitter. Error bars of ±1 S.D. are shown to illustrate deviations from linearity are not strictly due to random chance. **C.** When weighted by the total output spike count, information efficiency peaks at 15 ms of jitter and then decreases for inputs with smaller amounts of jitter.

gains while decreases in synchrony rapidly decrease the estimator precision. As the Fisher information is directly related to the local slope of the tuning curve this qualitative observation was unaffected, in a relative sense, by the discretization of the tuning curve. The raw information decreased approximately linearly with increasing minimum jitter as shown in Figure 3.7B (error bars are ±1 S.D.). However, as we will show the degree to which this is not linear has important implications for the efficiency of information transmission by the cortical neuron.

From these results, we naively assumed that a strategy that absolutely increased synchrony would always be best as it would always result in increasing stimulus information. As has been noted in other models which bear some similarities to our own (Wang et al., 2010a), there is a metabolic cost to increasing firing rate which can affect the efficiency of some information representations relative to others. In this case, as shown in Figure 3.7C, when we normalize the absolute amount of information by the number of cortical spikes, it becomes clear that the peak in transmission efficiency occurred at around 15 ms of thalamic jitter, and a quadratic fit had a peak at 16 ms with a clear decrease in information efficiency away from this peak. In previous

studies (Butts et al., 2007, Desbordes et al., 2008, 2010) we identified that pairwise LGN synchrony in response to natural scenes tends to be from 10 to 20 ms as measured by our scale. As noted, this result was consistent across all simultaneously recorded neurons when these neurons were used as sources for single-trial spike times. A few neurons maintained this quadratic relationship between information transmission efficiency and input synchronization at a peak efficiency closer to 25 ms of timing jitter, slightly lower than expected. These results indicate that populations in the LGN are uniquely arranged to be effectively synchronized by a preferred orientation. This synchronization allows information transmission to be more efficient without sacrificing precision in estimating orientation.

3.3.6 Tuning Width Invariance.

The results presented so far have demonstrated that information efficiency saturates at levels of minimum timing jitter between 10 and 20 ms, without addressing the effect of tuning width. It is clear from existing literature that there is a wide range of tuning widths that are typically measured in neurons in visual cortex (Anderson, 2000, Carandini and Ferster, 2000, Miller and Troyer, 2002, Rose and Blakemore, 1974, Gizzi et al., 1990) and these changes are reflected in the width of $\sigma_S(\theta)$ and thus the width of the tuning curve. To investigate the effect of changes in just tuning width we modulated both the minimum timing jitter as well as the tuning width, with the results shown in Figure 3.8. From 4.1 to 30.8 degrees (HWHH; maroon to light blue dots in Figure 3.8), which covers the rough range one could expect tuning width to vary, it is clear that the normalized information per spike (IPS) has approximately the same pattern regardless of tuning width. We show normalized information per spike because Fisher information is directly related to the slope of the curve, higher slopes monotonically lead to higher absolute levels of information and as such 4.2 degree and 30.8 degree tuning widths have an order of magnitude

difference in their absolute amount of information. The relationship between tuning width and information efficiency is made clearer in the breakouts in Figure 3.8B for each individual tuning width; with the exception of extremely narrow tuning widths, as the tuning width increases the optimal level of minimum jitter increases but still stays in the range of 10-20 ms. Narrow tuning curves fail to saturate information per spike because very narrow tuning curves effectively contain information about a very small range of orientations and the amount of information is directly related to the diference between baseline and peak firing rates. As an example consider a tuning curve that goes from baseline firing rate to peak firing rate in the span of 2 or 3 degrees (a very narrow tuning curve). In this case higher peak firing rates have a very pronounced affect on the overall amount of information. Since lower jitter always provides higher peak firing rates, narrower tuning curves are always most efficacious at extremely low amounts of jitter. We thus see that the results are valid for a range of primary visual cortex neurons so long as they have tuning widths that are within physiologically measured ranges.

3.4 Discussion

In this work we investigated the role of stimulus-driven synchrony in thalamic populations in the emergence of feature selectivity in primary visual cortex. The complete understanding of this role requires observation of entire thalamic sub-populations that are convergent onto single cortical neurons. Since these populations are too large to record electrophysiologically using current experimental methodologies, we synthesized representative populations from experimental data by randomly choosing recorded trials of neurons from which we could record, when obeying anatomical rules of thalamocortical connectivity (Jin et al., 2011) (also see below). These populations had an amount of stimulus-driven synchronization that was a direct function of the orientation of a drifting grating stimulus. These synthesized populations allowed us to

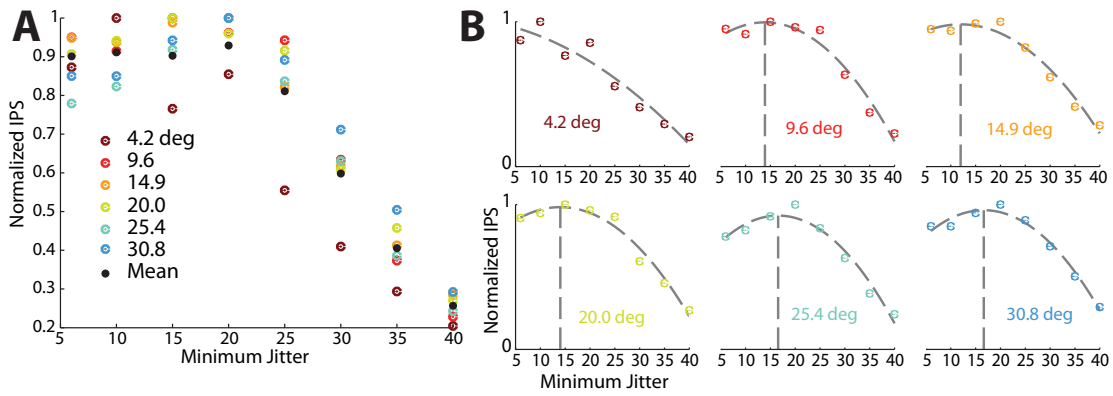

Figure 3.8. Quadratic efficiency is relatively invariant to tuning width. **A.** When taking the mean (normalized) efficiency curve across the spectrum of reasonable spike count tuning widths (HWHH, degrees), the arragement of optimal efficiencies does not appear to be patterened in any particular way. **B.** When broken out into individual efficiency curves we see that for each tuning width a quadratic polynomial still remains the best fit for most tuning widths. At pathologically narrow tuning curves, we see that higher synchronization is indeed absolutely preferable. We also note that the sigmoid fit to mean data arises, in part, because the peak of the polynomials are distributed over a range and the mean of them produces a roughly constant function below 25 ms of timing jitter.

systematically modulate the underlying spike timing synchrony to investigate the way in which different levels of synchronization affect information transmission. Through a biophysically inspired integrate and fire model that simulates cortical responses, we estimated the resultant cortical orientation selectivity and the corresponding information conveyed about visual stimulus orientation by the cortical response. Ultimately we found that the level of synchronization of the input population had a nonlinear effect on the resulting information contained in the cortical response; higher levels of synchrony led to higher levels of information, but at the expense of a nonlinear increase in firing rate. When taking into account the potential cost of increased firing rate, we found that the most efficient transmission of information was at a level of thalamic synchrony in the range of 10 to 20 ms.

3.4.1 Roles of Synchronization in Neural Responses to Visual Stimuli

It is important to note that the synchronization of neurons has been widely studied in a number of different contexts. Notably, synchronization of neurons across cortical columns has been previously reported in the visual cortex, proposed as a means to form relationships across regions of the visual field (Gray et al., 1989). Additionally, in the context of convergence and divergence of retinal afferents projecting to the LGN, precise correlations have been observed across geniculate neurons that were present in the absence of stimulus driven correlations, and were attributed to the projections of common retinal ganglion cell inputs (Alonso et al., 1996). In contrast, the current study (and previous studies from our group (Desbordes et al., 2008, Stanley et al., 2012)) specifically examines the role of stimulus driven synchronization/correlation of neuronal firing in the visual thalamus. Our previous investigations have shown that many neurons in the LGN do not exhibit appreciable noise correlations (Desbordes et al., 2008). The focus here is thus on the relationship between the visual input and the resultant synchronization of firing activity across geniculate ensembles, a requisite for robust activation of the downstream cortical neurons to which they project. In the most general case, however, as described in Gray et al. (Gray et al., 1989), the propagation of neuronal signals would involve a combination or interaction between the synchronization due to ongoing spontaneous activity and the stimulus-driven synchronization due to coordinated activation of neurons sharing the same topology and feature selectivity. Such a "from-any-source" view of synchronization carries with it the possibility that neurons with receptive fields from disparate regions of the visual field could be synchronized by spatially correlated stimuli. For example two very spatially distant LGN neurons could be simultaneously activated by either two unrelated objects or one very long bar of light; synchronization due to these origins are not considered in this model. It is important to note that we explicitly consider only recordings from spatially localized populations, as widely-spaced LGN units do

not converge at the same cortical target.

3.4.2 Comparison to Existing Models of Orientation Selectivity

The emergence of orientation selectivity in primary visual cortex is perhaps the most well-studied example of cortical computation to date. As a result, there have been a large number of modeling studies seeking to capture the mechanistic explanation for the primary observation of orientation selectivity, and also to capture a number of related, and more complex functional properties (e.g. contrast invariant orientation tuning, cross-orientation suppression, etc.). Given that there is little if any dispute as to the role of direct feed-forward geniculate input to cortical layer 4 in establishing the basic orientation preference for cortical neurons, models of orientation selectivity have invariably been constructed around a backbone of thalamic input. Although the model from Ringach introduced structured synaptic weightings and connectivity probabilities of thalamic inputs to cortex as a key model element (Ringach, 2004), the majority of other models assume relatively simple feedforward excitation structure and differ primarily in the relative strengths of the feedforward or intracortical inhibition (Finn et al., 2007, Anderson et al., 2001, Priebe and Ferster, 2005, Palmer and Miller, 2007, Kayser and Miller, 2002, McLaughlin et al., 2000, Somers et al., 1995). A specific limitation of most of these previous models is that they explicitly do not directly involve electrophysiological data as thalamic input. For example, one class of models use simulations of thalamic or retinal responses based on the stereotypical difference-of-Gaussians representation of center-surround receptive fields (Palmer and Miller, 2007, Kayser and Miller, 2002, McLaughlin et al., 2000, Ringach, 2004, Somers et al., 1995), while others rely on assumed or derived cortical conductances or membrane potential but not on actual thalamic input (Finn et al., 2007, Anderson et al., 2001, Priebe and Ferster, 2005). The large majority of previously published models also assume that sinusoidal inputs (i.e. drifting gratings) elicit sinusoidal

thalamic responses and that the cortical membrane potential itself is perfectly sinusoidally modulated (as in (Anderson, 2000) or (Finn et al., 2007, Palmer and Miller, 2007)). Dating back to the early 1980s there was the observation that drifting sinusoidal gratings produced asymmetric LGN response PSTHs (i.e. a sharp peak at the onset of the stimulus followed by a long tail of decaying response) (Derrington and Fuchs, 1979, Hicks et al., 1983, Saul and Humphrey, 1990) and more recently we have directly analyzed the effects of this synchrony in the context of cortical orientation and direction selectivity (Stanley et al., 2012). We assert that the precise timing and stimulus-driven synchronization of thalamic inputs serves a prominent role in the thalamocortical circuit and in the emergence of cortical feature selectivity.

Most, if not all, existing models designed to capture the mechanism behind cortical orientation selectivity rely on spatial arrangements of projecting thalamic inputs that in some cases exceed those observed experimentally (Jin et al., 2011). More specifically, the relevant measure for thalamic input is the aspect ratio of the scatter of thalamic receptive fields that form the input to a single cortical layer 4 neuron. Recently, Jin et al. experimentally observed thalamic clusters and showed that the thalamic input to cortical orientation columns has receptive fields that are highly overlapped (Jin et al., 2011). Because the scatter of the thalamic receptive fields covers 2.5 receptive field centers in visual space, the average layer 4 cortical neuron should have a maximum aspect ratio of 2.5:1. The thalamocortical model from Somers et al. was built on an aspect ratio of 3:1 (Somers et al., 1995), whereas the model from McLaughlin et al. was built on an aspect ratio of 4:1 (McLaughlin et al., 2000). Similarly large aspect ratios are apparent from the Kayser et. al. model and Finn et. al. models, with ratios approximately 6:1 and 2.5:1 respectively (Kayser and Miller, 2002, Finn et al., 2007). It is clearly the case that inhibitory mechanisms play a significant role in the shaping of the cortical feature selectivity (Ferster and Miller, 2000), and would only serve to further refine the selectivity established by the direct

feedforward thalamic input shown here. Many of the above-mentioned models differ from our presentation here in that they include OFF-center sub-populations in the thalamic population, most commonly offset from the ON-center population as would be implied by the common Gabor-type simple cell receptive field. To keep the model relatively straightforward and simple, we have chosen to focus on just ON-center populations.

The majority of existing models were optimized to explain extra-classical effects of cortical receptive fields with a particular focus on the contrast invariance of cortical tuning width and as such constructed mechanisms specific to this issue. Specifically, it has been widely observed that although peak cortical firing rates are strongly dependent upon stimulus contrast, cortical orientation tuning is largely invariant to stimulus contrast (for review, see (Ferster and Miller, 2000)). This observation called into question the purely feedforward model of cortical orientation selectivity (Ferster and Miller, 2000). Subsequent models augmented the feedforward thalamic input with inhibitory feedforward connections (Finn et al., 2007) or cortico-cortico inhibition (Kayser and Miller, 2002) or some combination (Somers et al., 1995, Ferster and Miller, 2000). We have previously shown that thalamic synchrony is largely unaffected by stimulus contrast (Desbordes et al., 2008), and the cortical tuning based on thalamic synchrony is also contrast invariant. The model we have proposed here thus potentially demonstrates a completely feed-forward explanation for contrast invariance. For a fixed minimum jitter amount, as the underlying LGN firing rates across the entire population are modulated by changes in the stimulus contrast, the peak induced firing in the cortical neuron rises and falls. Since the changes in LGN firing are correlated across the LGN population, the synchrony across such a population (with particularly focus on the relationship between stimulus orientation and the synchrony) remains unchanged as a function of stimulus contrast. As demonstrated in Figure 3.5B for the span of biophysical levels of preferred orientation population

synchrony (∼5 to 20 ms), the tuning width of the cortical neuron does not change, indicating that changes in the degree of underlying synchrony do not change the tuning properties. Although the results are not presented here directly, the combination of past and present results suggest that changes in the LGN population response (i.e. the population becomes less active in general) lead to a decreased or increased peak cortical response but the tuning curve widths will be invariant to stimulus contrast.

3.4.3 Fisher Information in Neural Populations

We used Fisher information as a measure of the efficacy of cortical neurons in representing stimulus features (orientation) in response to changes in the synchrony of an input population. Specifically, we used the peak Fisher information irrespective of the orientation at which the peak occurs. Contrary to previous investigations (Xie, 2002, Butts and Goldman, 2006, Yarrow et al., 2012) in which the absolute value of the Fisher information was used as an important measure of the performance of neural populations, here we sought to capture the relative effects of varying degrees of thalamic synchrony on the information conveyed by a single recipient cortical neuron target. In this case, we assumed that the Cramér-Rao lower bound need not be met and that whatever bias causing deviations from the lower bound is consistent across all simulation conditions. We ensure this by using the same input data and model structure for all conditions so that we can compare relative levels of information across different synchrony conditions for a single neuron. Although this is a simplification of the true amount of information (and indeed no single neuron will saturate this lower bound), in either case the absolute information was consistent with previous studies utilizing experimental cortical data. Yarrow et. al. (Yarrow et al., 2012) computed Fisher information for both real and simulated neural populations and found an information level which was approximately consistent with the findings presented here (see their Figure 4 as well as (Butts and Goldman, 2006) Figure 3, with axes in

(Yarrow et al., 2012) helping in the conversion from SSI bits to Fisher Information in units of deg^{-2}). This assumption ultimately only affects the reporting of estimator standard deviation (as in Figure 3.7A) which was not the primary result of the work.

The application of Fisher Information to cortical tuning curves has deeper roots in estimating cortical population response information transmission. Past work (Paradiso, 1988, Seung and Sompolinsky, 1993, Abbott and Dayan, 1999, Pouget et al., 1999, Zhang and Sejnowski, 1999) has in general used constructions where a collection of identical cortical neurons have preferred orientations that uniformly span the orientation spectrum (0 to 360 degrees). In this study we considered only a single neuron in the population. We claim, though, that results which demonstrate information in a single neuron at all stimulus orientations are fundamentally identical to results which demonstrate information in a population at a single orientation. As long as we assume every neuron in the cortical population is conditionally independent, for the questions we ask these two formulations are fundamentally interchangeable. As identified in (Seung and Sompolinsky, 1993) under the assumption that each cortical neuron in this population is independent, then at every stimulus orientation the overall Fisher information is

$$J(\theta) = \sum_{i=1}^{N} J_i(\theta)$$

Further, in the case that every neuron in the population is also assumed to be identical in response properties, then we can modify the above to read (for any choice of i)

$$J(\theta) = \sum_{k=1}^{360} J_i(\theta_k)$$

It is clear though that not all cortical tuning curves are identical and the absolute amount of information is strongly negatively correlated with tuning width. Using this fact as inspiration, we show in Figure 3.8 that the optimally efficient level of input timing jitter is widely insensitive to the tuning width of the cortical neuron. In this case, even if a cortical population is composed of non-identical independent neurons,

each neuron, as well as the population as a whole, will be optimally efficient as long as the thalamic input is synchronous to the 10-20 ms level (thus implying we need no longer assume neurons within the population have identical, but shifted, tuning curves). If we further consider the effects of correlated variability, as in (Abbott and Dayan, 1999), then we can no longer assume the units are independent. Regardless of whether the correlated variability increases or decreases the absolute amount of information (and neither is guaranteed), correlated variability would raise or lower the response rate of the individual neurons in a coordinated manner. Since again our metric is one of relative comparisons, the results presented here are expected to be invariant to correlated variability in the sense that the efficiency of any single neuron may decrease, the peak efficiency will still occur between 10-20 ms (which would still be true for all neurons in the cortical population). Thus our findings directly translate to cortical populations regardless of the independence and homogeneity of tuning properties of the component neurons.

3.4.4 Timescales of Natural Vision

In previous studies of timing precision of individual thalamic neurons (Butts et al., 2007) and across thalamic pairs (Desbordes et al., 2008) in response to natural scenes, we have reported characteristic timescales on the order of 10-20ms. In these previous studies, measures were taken across long segments of natural scene movies, representing the aggregate of instantaneous firing events whose timing precision clearly varies on an event-by-event basis (Desbordes et al., 2010, Butts et al., 2010). The instantaneous synchronization of firing activity across a sub-population of neurons in the context of natural scenes is undoubtedly a complex function of the local properties of the scene, including spatial frequency, temporal frequency, and orientation of the local spatial structure. It is thus the case that the 10-20 ms average timescale reflects a distribution of synchronous events, spanning from synchrony on just a few milliseconds

to more asynchronous firing over a timescale of 10s of milliseconds, unlikely to drive the cortical target. Here, we report that in the context of the modulation of thalamic synchrony through visual stimulus orientation with drifting sinusoidal gratings, the most efficient level of thalamic synchrony in conveying relevant information to cortex is in the 10-20 ms range. This means that, on average, amongst natural scenes and all their various features, the thalamic neural response is tuned to maximize the efficiency of information transfer to the cortex (similar to (Wang et al., 2010b)). As we have investigated only the effects of orientation changes on synchronization and feature selectivity, we expect that this result implies that information efficiency will be similarly optimized for other visual features such as spatial and temporal frequency. Furthermore, it is possible that synchronization optimizes information transmission in entirely different sensory systems, given previous findings in the somatosensory system (Stanley, 2013).

CHAPTER IV

NATURAL STIMULI AND COMPLEX MOTION

4.1 Introduction

4.1.1 Visual Scenes: Natural or Artificial?

For many decades, neural activity in the visual system was viewed through the lens of artificial visual stimuli, in an attempt to dramatically reduce the available stimulus space. From the groundbreaking work of Hubel and Weisel (Hubel and Wiesel, 1962) and others (Campbell et al., 1969, De Valois et al., 1982b, Bradley et al., 1987, Shapley and Lennie, 1985, Priebe et al., 2006), we have learned a tremendous deal about how neurons in the primary visual cortex actually integrate basic properties of the outside world and respond to these properties. As discussed in previous literature (Rust and Movshon, 2005), use of these stimuli is effective and efficient because it allows the experimenter to control explicit parameters as opposed to the comparatively chaotic realm of natural scenes. The principle behind reducing the stimulus space in this manner is that by understanding the constituent elements of a scene, even a complex natural scene, and how the elements drive visual neurons, we can build up an expected representation of that scene through these individual elements. This perspective requires the ability to grind up such stimuli into their component elements.

It is not clear, though, what actually constitutes a natural scene despite this decomposition perspective, as the relationship between continually changing parameters (within space) and fixed parameters in artificial stimuli such as the classic drifting sinusoidal grating and the corresponding neural responses is not clear. Natural scenes are usually defined as any visual stimuli that have not been manually created and they are often drawn from databases of images that contain pictures and movies from

the outside world. At some level, a natural scene can be described as paraphrased from Former U.S. Supreme Court Justice Potter Stewart; "I know it when I see it" (Jacobellis v. Ohio 378 U.S. 184 (1964)). As our use and understanding of natural scenes has grown, the most common description of them has been a statistical description of their frequency content and the associated power (Field, 1987, Reinagel and Zador, 1999, Simoncelli and Olshausen, 2001). These thorough analyses found that luminance is relatively smooth in natural scenes and highly correlated within the range of a few pixels; the measured spatial frequency power spectrum has been shown across multitudes of natural scene images to fall off roughly following the curve $1/f^2$ which indicates a strong preference for low spatial frequency content. Similar types of analysis have looked at color spectra in natural scenes, different types of motion boundaries (Roth and Black, 2007), and even the number and types of objects present in natural scenes (Li and Perona, 2005, Greene, 2013). After all this analysis of natural scenes, however, when such stimuli are used in experiments they tend to take the form of the same chaotic stimuli we receive every day. There are few attempts at generalizing the understanding of natural scenes to more contrived contexts in which the scene is natural-like yet still exists in a way that can be readily interpreted and correlated with observed neural responses.

Reductions of natural scenes have been used in some contexts, like navigation in the case of Saleem et al. (2013) or in rotational representations of simpler patterns (Rosenberg et al., 2013, Liu et al., 2004). These studies revealed important neural coding parameters, yet none explored the consequences of these stimuli in light of the kinds of feature selectivity represented in the central visual pathway. This is particularly important given the kind of critical view some take of the usefulness of natural scenes for retinal, thalamic, and primary visual cortex contexts (Rust and Movshon, 2005). Fundamentally we typically ask questions about individual stimulus properties and how they may combine to create a neural response, but given

the underlying importance of motion within the scene it is necessary to peer deeper and look at how more complex motion (which is represented in parameter changes) modulates the response of neurons. Currently this modulatory affect, particularly in primary visual cortex, is unclear.

4.1.2 The Three-Dimensional World

An important aspect of natural scenes is that we obviously live in a three dimensional world. We perceive the world, however, through a fundamentally two-dimensional representation. When light hits retina, it is projected onto a two dimensional sheet of retinal ganglion cells which preserve the relative locations between objects. One natural side effect of this is that scaling becomes very important; two identical objects will appear to be very different sizes if they are placed at different distances from the observer (Kandel et al., 2000). This effect is demonstrated for local spatial frequency as shown in Figure 4.1A. For an observer that is centered on a tilted sinusoidal grating, the change in apparent local spatial frequency is due to the change in incidence angle of a single period of the sinusoidal grating. At each end of the texture, the change in incidence angle can be significant. When projected through to the retina both periods cover the same physical distance b but the retinal distance would be much smaller (compressed) for smaller incidence angles or expanded for larger incidence angles. Spatial frequency thus changes as a function of image location once projected. This is seen in Figure 4.1B, using a novel generated image. In this it is clear that along the red line luminance values have a differing local spatial frequency depending on location within the image.

Perceptually, we resolve these incongruities because although we only receive 2D information from the outside world, the offset between our eyes means we receive two slightly different 2D representations of the same 3D world, and this generates enough new information that we can completely recover depth information (Purves

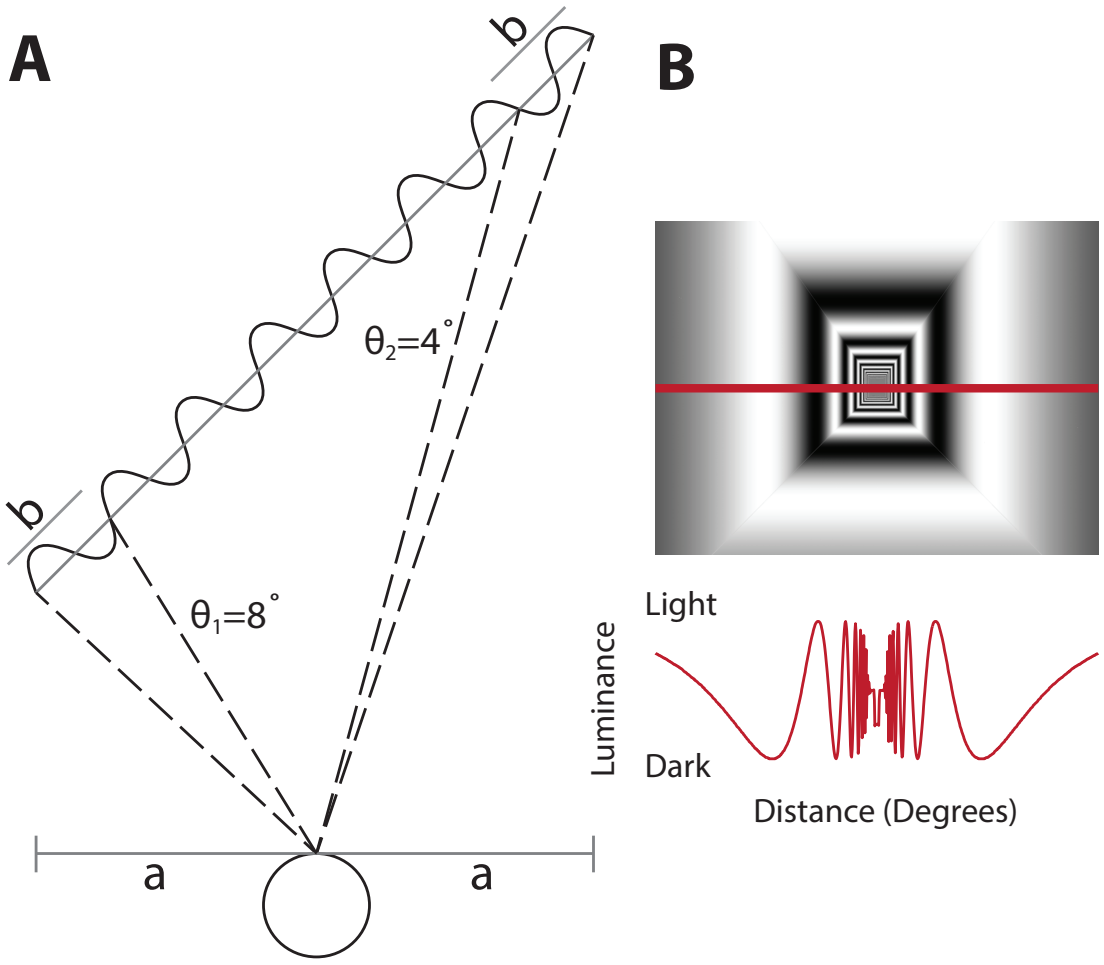

Figure 4.1. Effect of Perspective on Perceived Spatial Frequency. A. For a viewer centered on a rotated sinusoidal grating, periods of the grating subtend different angles of the visual field based on their distance to the observer. When projected to a two-dimensional representation these different angles translate to compressed or expanded representations of identical physical space (b, 1 period) resulting in spatial changes in local spatial frequency. **B.** This is shown using a novel stimulus which has been generated with perspective. Along the red line in the image (top) the local spatial frequency changes as a function of location in the image, reflected in a plot of image luminance along that line (bottom).

et al., 2004). On a continuum, this decrease in size results in a visual perspective in which the sizes of objects decrease in a continuum until they reach a singular point on the horizon (Gibson, 1979). Computationally this means that classical stimulus properties like spatial and temporal frequency are based on the distance of the object as well as the intrinsic object properties, although the resulting perception of motion and speed is unaffected. This is another element of the natural scene; where classical stimuli evoke a flat view that is typically devoid of the artifacts of perspective and a three dimensional world, all truly natural scenes include this as an important element. In some scenes, the configuration of the objects do not provide a horizon and thus the perspective effect is limited (for example, a picture of a room) while others evoke strong perspective (for example, looking down 42nd Street towards the Hudson River in New York City). As neurons in various areas of the visual pathway have been found to evoke depth-related responses from both texture and shading cues (Liu et al., 2004, Rosenberg et al., 2013), even if we capture neural responses when presenting two dimensional images to a single eye (and thus eliminating the possibility of any binocular depth information existing), it is not clear how the responses to classical grating stimuli would differ from those in which the grating has been forced to have a perspective but otherwise identical properties.

4.1.3 Interface Between Natural Scenes and Motion

In previous work we have demonstrated that thalamic populations appear to give rise, through synchronization, to efficient representations of stimulus orientation and directional selectivity (Kelly et al., 2014, Stanley et al., 2012). This finding was hypothesized to generalize to all feature selectivity but this selectivity was generated under conditions of artificial stimuli. It is natural to ask, then, whether or not such commonly observed feature selectivity, often represented through 1-dimensional tuning curves or 2-dimensional tuning surfaces, actually captures the response properties

of neurons under natural conditions. We hypothesize that when the visual system is presented with natural-like scenes, i.e. those that mirror some of the phenomena in full natural scenes (such as Hollywood movies, cat-cam sequences, or natural scene movie sets) but with significantly reduced visual complexity, the observed responses will different significantly from what would be expected by the simplest and most classical tuning-curve driven models which use a limited number of tuning dimensions.

Neural populations are clearly important to our concept of feature selectivity and along with receptive field representations, this feature selectivity represents the sum of what we know about how visual neurons represent the outside world. Feature selectivity and tuning are so important to our understanding of these visual neurons that much of our theoretical and experimental analysis is done to predict its cause and importance within information transmission (Seelig and Jayaraman, 2013, Neri, 2014, Ruiz and Paradiso, 2012, Grunewald and Skoumbourdis, 2004, Cumming and Parker, 1999, Sompolinsky and Shapley, 1997, Bridge and Cumming, 2001, Shapley and Lennie, 1985, Wilent and Contreras, 2005, Perrone, 2004, Smith et al., 2006, Liu et al., 2004). What is more at stake in this question, however, is how this feature selectivity transmits information about the motion in the stimulus rather than how accurately it lets us pinpoint orientation or spatial frequency features. Given that these two features are inextricably linked, it is important to know how well feature selectivity measured under artificial conditions corresponds to the responses of the same populations under natural stimulus conditions. Knowledge of this link provides a tremendous amount of clarity to how populations and single neurons transmit and represent this stimulus motion information.

To address this question of the generalizability of tuning curve approximations of neural responses to motion, we generated a unique natural-like stimulus. It is called the "sinusoidal hallway" and uses sinusoidal textures mapped to the walls of a hallway

that has a far-away convergence point. By walking down the hallway and actively exploring the space, we can manipulate features like perspective warping and motion boundaries while monitoring how well we can both encode single-neuron responses and decode the stimulus properties from a large population of neurons. The results indicate that our understanding of feature selectivity is currently incomplete and that our understanding of tuning lacks sufficient dimensionality to capture the breadth of responses to natural scenes. This failure could go a significant way to explaining the errors consistently observed in natural scene response prediction, particularly in primary visual cortex using tuning-curve based approximations (Weliky et al., 2003). Despite these challenges, decoding of different sequences within these two stimuli remained highly effective across an entire population of V1 neurons, indicating that information about the complex motion within these simple stimuli is embedded in the whole of the population. To account for the observed responses, a model that nonlinearly accounts for natural-like features could also unlock this information and allow us to generate perspective and motion information even from monocularly driven primary visual cortex populations. Although the results are not shown here, preliminary investigations using receptive-field based LN models appeared to fare much better at generating accurate encoding predictions. This is likely due to the fact that a very accurate receptive field model embodies all manners of feature selectivity and nonlinearities present in a single V1 neuron, rather than just a handful explicitly measured properties.

4.2 Methods

4.2.1 Electrophysiology

Surgeries were performed identically to those in Chapter 2 and as has been discussed in previous publications (Stanley et al., 2012, Kelly et al., 2014). Neural recordings were captured from primary visual cortex of anesthetized cats. Electrophysiological

recordings were made using dual Neuronexus 32-electrode probes, one placed perpendicular to the cortex and a second placed approximately parallel. The perpendicular probe was aimed to traverse down an orientation column while the parallel probe was aimed to cross multiple columns. Each electrode site was separated by 100 um for a total length of 3.2 mm. Across all penetrations a total of 868 channels were recorded from across both ipsilateral and contralateral eyes.

4.2.2 Visual Stimuli

Visual stimuli were presented on a CRT monitor at 120 Hz with presented images gamma corrected to ensure a linear representation of the entire white to black range of the monitor. All stimuli were presented using the PsychToolbox toolbox for MATLAB (available from https://www.psychtoolbox.org/HomePage). Receptive fields were mapped using sparse light and dark noise to individually capture light- and dark-driven subfields with reverse correlation capturing the resulting receptive field maps. Sparse noise uses a whole-field constant luminance and individually triggers stimulus pixels to an opposite luminance; a light-field stimulus puts black pixels on a white background (Jones and Palmer, 1987). Individually captured subfields tended to occupy the same space but often with different radii of each subfield, but due to the relatively large pixel size used for the sparse noise signals the receptive field size is potentially over-estimated. Spatial frequency and orientation tuning properties were simultaneously measured using a series of drifting sinusoidal gratings that swept across a parameter space of both spatial frequency and orientation. Orientation ranged from 0 to 337.5 in 22.5 degree increments (16 total orientation) and spatial frequency ranged from 0.03 to 3 cycles per degree (10 total increments logarithmically spaced) with 5 repeats for each orientation-spatial frequency pairing. The convention is that a rightward traveling vertical grating was 0 degrees, a downward traveling horizontal grating was 90 degrees, and so on. The spatial resolution for all stimuli

was 0.056 degrees per pixel. The temporal frequency for all drifting gratings was 5 Hz and the hallway texture gratings drifted at approximately 3 Hz. All stimuli were presented at 100% contrast.

4.2.3 Hallway Stimulus

Using the 3D design program Blender (http://www.blender.org/) we generated the hallway stimulus. Textures for use in the hallway were created in MATLAB (Mathworks). The texture was created with a spatial frequency of 0.15 cycles per degree to provide a spatial frequency range (when projected) within physiological range for V1 neurons (Movshon et al., 1978) and a temporal frequency of 3 Hz for the same reasons. The camera captured a typical 35mm depth of field in order to reproduce the effects as we might typically see them in regular photography and film. After a trial run, we modified the stimulus to have the perception of being visibly infinite by creating it to be incredibly long relative to how fast we travel down the hallway. This design allowed us to maximize the visually useful areas and created the most realistic view of a sinusoidal grating possible. To provide luminance consistency shadows were completely disabled in the modeling. After 3D modeling in Blender and rendered to 2D, stimulus image sequences were analyzed to verify temporal and spatial frequency properties. The sinusoidal hallway stimulus was presented both with a forward presentation (walking forward) and a backwards presentation (walking backwards).

Although the texture that defines the walls within this hallway was created to exacting standards (i.e. an exact spatial and temporal frequency), manipulation within the Blender program makes maintaining this information difficult. For the sake of aesthetics and artistic manipulation the software does not have an interface to determine the exact parameters used to project the 3D world down to 2D. Using Blender to design the stimulus limits access to the distance and rotation information (see Figure 4.1) necessary to algorithmically determine local spatial frequency. Alternatively we

could have harnessed the algebra of Blender's orthographic projection,

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \boldsymbol{P} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

and the relationship between (x, y, z) pairs and their projected (x', y', z') pairs would allow an algorithmic determination of the resultant rendered spatial frequency as a function of space. In practice, owing partially to the difficulty of obtaining useful coordinates and projection matrix representations from Blender itself, it is easier simply to estimate spatial frequency as a function of space for the output rendered image sequences. In this case (since luminance is bounded and reliable) the peaks in the luminance curve that cuts across the sinusoidal function are calculated for each frame. From each peak we estimate the distance to ± 0.25 cycles (corresponding to the luminance value 128) and use that half-cycle distance to compute the local spatial frequency. Since the 3 Hz drifting is sampled at 120 Hz, the visual space is sampled heavily making it very easy to fit a function to the spatial frequency relative to location within the stimulus.

4.2.4 Data Analysis

4.2.4.1 Receptive Field

Individual ON and OFF subfields were mapped using sequences of dark and light sparse noise stimuli. Since we chose not to perform any spike sorting we could not determine simple versus complex cell classifications. Using a standard reverse correlation technique the recorded spike times were used to generate the receptive field kernels. The kernels were mapped to a 50% (half height) kernel and then fit with an ellipse to obtain a smooth representation of the receptive field. The size of each RF was calculated as the average radius between the ON and OFF subfields; in nearly every case the ON and OFF subfields were located on top of each other so the average

radius captures the effective receptive field size. The average receptive field size was 2.58 ± 1.15 degrees (N=868).

4.2.4.2 Orientation Tuning

Orientation tuning of the V1 neural activity was measured from the pairwise orientation-spatial frequency tuning surface, calculating the orientation tuning curve at the strongest responding spatial frequency. The tuning curve captures the mean response to a presented stimulus (800 ms) and was fit with a two-Gaussian model after shifting the preferred orientation to be at 90 degrees

$$f(\theta) = a_1 * \exp\left(-\frac{(\theta - 90)^2}{2b_1^2}\right) + a_2 * \exp\left(-\frac{(\theta - 270)^2}{2b_2^2}\right) + c$$

After accurately capturing the statistics of preferred orientations, we chose to shift the preferred orientation so that fitting the tuning curves would not require the use of more complicated circular Gaussian models such as a von Mises function (as in Wissig and Kohn (2012)). Tuning curves were only fit to get an accurate estimate of tuning width, and this fit is independent of the preferred orientation. The HWHH tuning of each channel was computed for each of the two Gaussians based upon the standard deviation, where $HWHH = 2.335 * \sigma$. The average HWHH tuning width was 40.3 ± 19.4 degrees (N = 868), with a minimum of 16.5 degrees.

Directional Index (computed as in Stanley et al. (2012)) was also computed for each orientation tuning curve, using the equation

$$DI = 1 - \frac{R_{\theta_0+180}}{R_{\theta_0}}$$

where θ_0 is the preferred orientation, R_{θ_0} is the mean firing rate at the preferred orientation and R_{θ_0+180} is the mean firing rate 180 degrees off the preferred orientation. A directional index value of 0 indicates that there is no directional selectivity and axially identical stimuli (for example vertical gratings that drift leftward or rightward) would elicit identical responses. A directional index value of 1 indicates that the

response is perfectly directionally selective. The average directional index was 0.47 $\pm$ 0.21 (N = 868), indicating a moderate level of directionality.

4.2.5 Tuning Function Based Modeling

Encoding of the responses to the sinusoidal hallway stimulus was modeled as a combination of an underlying sinusoidal modulation driven by local changes in luminance and an amplitude function driven by local changes in spatial frequency and orientation. Generally put

$$R_n(t) = A(\theta(t), \omega_s(t)) * L(x_n, y_n, t)$$

where (x_n, y_n) refers to the center of the receptive field of the nth neuron, $R_n(t)$ is the observed response PSTH, θ is stimulus orientation (defined either at each pixel within a receptive field area or at the center of the receptive field, depending on the chosen method), ω_s is the stimulus spatial frequency, and L is the stimulus luminance. As the underlying stimulus is modeled off of a sinusoidal grating, we used our combined spatial frequency and orientation response measurements to generate a "tuning surface" (opposed to a one-dimensional tuning function). This measures how well the orientation and spatial frequency pairs of parameters drive the neuron and thus allows for a rough prediction of the mean response of a neuron to local parameters within our hallway stimulus. This local prediction from the tuning surface forms our estimate of $A(\theta, \omega_s)$ and we generated this value from two possible methods: the first where θ and ω_s depend only on the properties present at the center of the receptive field and the second where it is a weighted average across the entire receptive field space. In the first method, we look only at the parameter values present in the center of the receptive field (across both dark- and light- subfields) for a given stimulus frame. This unique pairing of stimulus properties uses the tuning surface as a mapping function to return an expected maximum firing rate for that pair of stimulus properties for that particular stimulus frame. In the second method, we

perform this lookup for every pixel within the receptive field space independently and the resulting expected maximum firing rate for every pixel is averaged together to approximate an expected firing rate for the neuron (receptive field) as a whole. An alternative would have been to weight the per-pixel firing rate expectation by the local strength of the receptive field for that pixel which allows for spatial variations in spatial frequency and orientation. Ultimately this method is practically equivalent to looking at just the center parameters, due to the fact that the receptive fields are generally symmetrical in size and distribution of strength. Regardless of the method chosen to determine A, the process is repeated for every frame in the stimulus to give a time-varying approximation of the amplitude of the stimulus modulations. As above, these are multiplied by the luminance in the center of the receptive field (regardless of the amplitude method) to generate a time-varying approximation of the instantaneous firing rate.

The quality of this encoding estimate is evaluated through the correlation coefficient between this prediction and the observed PSTH across all trials. To account for the underlying reliability of the recorded PSTH, this correlation is normalized by the split-in-two correlation of the PSTH (i.e. the correlation of half of the recorded trials with the other half of trials, in randomized groups, to estimate the reliability). This gives us a metric similar to the metric of percentage of variance explained (Millard et al., 2013), used to capture the efficacy of model predictions. Computations are only performed for PSTHs with a reliability greater than 0.5. This value is bounded by a minimum of 0 and a maximum of 1, with accurate estimates of the response encoding being considered as percent of variance explained values greater than 0.6. Given that the luminance at the center of the receptive field does not account for the temporal properties of the receptive field kernel, the represented prediction coefficient is the maximum across any latency.

4.2.6 Hallway Decoding Analysis

Decoding of the sinusoidal hallway was performed using maximum likelihood estimates using both a classification method as well as a true parameter decoding method. For classification, response distributions for two distinct segments (straight; from 0 to 1 seconds and turned; from 2.3 to 3.2 seconds) of the hallway stimulus across both the forward and reverse presentations were calculated using half of the available trials. From the remaining half of trials a random trial was chosen and the mean response during a random segment and direction was chosen. The maximum likelihood classifier

$$\hat{\phi} = \arg\max_{\phi \in \Phi} P(R_n|\phi)$$

where $\Phi = \{FS, FT, BS, BT\}$ (F - Forward, B - Backwards, S - Straight, T - Turned) found the most likely segment and direction to have resulted in the observed response. The population classifier is based on assuming that each channel is an independent observation of the stimulus and their maximum likelihood functions multiply.

For decoding of raw stimulus parameters the method is more complicated because the parameters change as a function of space. Again the stimulus is separated into two segments (straight and turned as above), but these two segments are no longer compared and only forwards presentations are considered. For each individual neuron, a log likelihood matrix is created which captures the likelihood that different pairings of spatial frequency and orientation resulted in the randomly chosen single-trial response from a segment - this matrix is fundamentally similar to the tuning surface but instead represents the probability that each of the parameter pairings had been presented. In this case, the training dataset comes from the tuning surface response distributions and the test data comes from single trial responses to the hallway stimulus, so there was no need to compartmentalize the data into two unique nonoverlapping sets. Since the underlying parameter changes are spatial though, this

log likelihood matrix is stored for every pixel within the receptive field of each neuron and in places where different neurons have overlapping receptive fields, the log likelihood matrices add, assuming conditional independence

$$\hat{\theta}(x,y), \hat{\omega_s}(x,y) = \arg\max_{\theta,\omega_s} \sum_{n=1}^{N} \log \mathrm{P}(R_n(y,x)|\boldsymbol{\theta}, \boldsymbol{\omega_s})$$

In this manner overlapping receptive fields help to multiply-sample the visual space in an attempt to generate more accurate estimations of the local stimulus parameters and allowing the estimates to vary on a per-pixel basis provides accuracy at a more fine-grained level than that presented by a single cortical receptive field.

4.2.7 Estimation of Response Correlations

To validate our assumption that the electrode channels we have recorded (as a stand-in for single neurons) are independent, we estimated the pair-wise correlations ($C_{i,j}$) between all simultaneously recorded neurons. Correlations between two neurons come in two forms: correlation caused by two neurons receiving identical sets of input (noise correlations, $C_{ij,N}$) and correlation caused by neurons receiving similar underlying stimuli (stimulus-driven correlation, $C_{ij,S}$). When stimuli are presented both of these types of correlation are present,

$$C_{ij} = C_{ij,N} + C_{ij,S}$$

True independence between the neurons in our population requires a low level of noise correlation. As we have little recorded activity under conditions of no stimuli, we had to estimate the noise correlations by subtracting an estimate of the stimulus-driven correlation from measurements of the full correlation

$$C_{ij,N} = C_{ij} - C_{ij,S}$$

The full correlation is calculated by computing the correlation between a given i, j pair of neurons using matched trials for each (Figure 4.10B). To measure the stimulus-driven correlation $C_{ij,S}$ we calculate the correlation between a given i, j pair of neurons

choosing a different random trial for each (Figure 4.10C). This trial-shuffled correlation captures the stimulus-driven correlations and eliminates any noise correlations (Desbordes et al., 2008, Ince et al., 2013, Briggs et al., 2013, Usrey et al., 2000).

4.3 Results

4.3.1 The Sinusoidal Hallway

To measure the effect of spatial nonlinearities inherent to natural scenes on the predictability of visual neural responses, we generated a natural-like visual stimulus and presented it while recording from multi-electrodes in the primary visual cortex of the anesthetized cat (see Methods). This stimulus is called the sinusoidal hallway as it has sinusoidal textures mapped to the walls of a three-dimensional hallway which the camera explores along a defined trajectory. This defined trajectory can move either forward or backwards (Figure 4.2A) while also implementing changes of the angle of the camera as if the observer's head was rotated to the left (Figure 4.2B).

As is shown in Figure 4.2C, the hallway stimulus contains 3 distinct segments, characterized by the angle of the camera within the scene. The first segment (within the first second of the stimulus) is straight in that the camera views the hallway head-on so that the center of the hallway is centered on the screen. Within this phase, spatial frequency and orientation, the primary descriptive properties of this stimulus, are fixed for any given point. The second segment (from 1 second to 2.3 seconds) is "dynamic" in that the head angle is actively changing. Although slow, stimulus parameters are changing between every frame in this segment. The final segment (from 2.3 seconds to 3.2 seconds) is the turned segment in which the head/camera has completed turning to look towards the left wall and holds that fixation. As in the straight phase, the stimulus properties are fixed for any given location during this segment. Example frames from each of the segments are shown in Figure 4.2D. Within each segment, the parameters also vary as a function of space such that

as you move away from the center of the image (either horizontally or vertically) spatial frequency continuously changes. There are also distinct boundaries between orthogonally orientated motion fields, the location of which changes as the head angle (i.e. location) changes.

4.3.2 Encoding of Hallway Responses

To determine how well the recorded neurons represented the spatial information within the novel natural-like sinusoidal hallway, we first attempted to model the encoding of the stimulus information. Encoding models provide the first clue as to how useful tuning properties are in predicting information flow in populations by verifying the generality of the measured tuning curves. We captured the mean response of each neuron to simultaneous variations of orientation and spatial frequency creating a tuning surface that represents all pair-wise parameter tuning responses. Using the tuning surface as a model of neural response, we assumed that local spatial frequency and orientation would guide the frame-by-frame firing rates observed by each neuron. Even in this simplified stimulus there are complex changes in local orientation and spatial frequency. An exaggerated representation of this can be seen in Figure 4.3, where different milestones in the stimulus can be seen projected onto an example orientation-spatial frequency tuning surface. At the start, local stimulus parameters have some particular value, here we have a low-frequency 180 degree orientation stimulus. As we rotate our head angle while moving through the hallway, for this example we slowly increase local spatial frequency until our view jumps to a different motion field with a different orientation. From there spatial frequency might increase significantly while keeping a fixed orientation until it hits the point at which the head angle ceases changing and eventually we reverse the process, ending where we started in the orientation-spatial frequency parameter space. While this is a magnified example of a typical trajectory, it makes it clear that response estimations have to account for

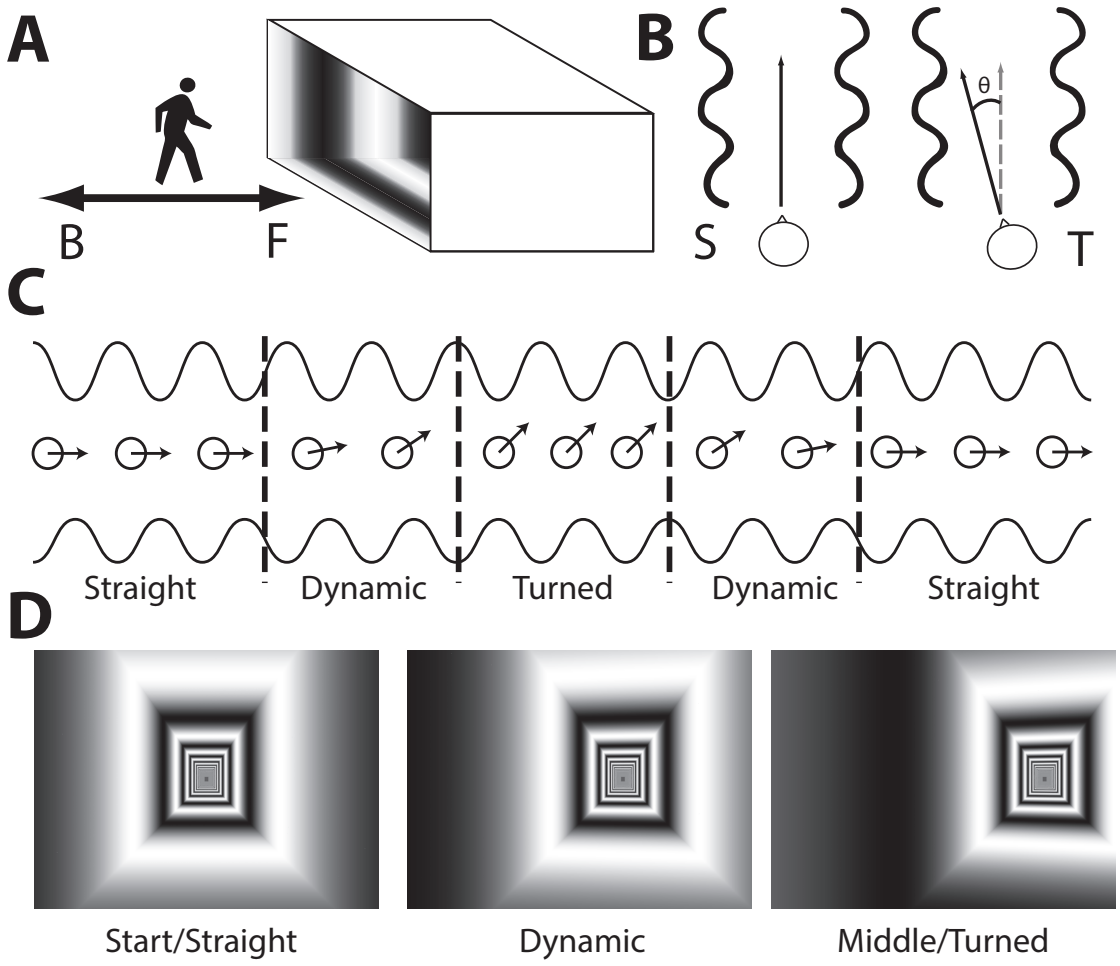

Figure 4.2. The Sinusoidal Hallway. A. The stimulus contains sinusoidal textures mapped to the wall of a hallway. The camera (emulating a person) can walk forwards or backwards. **B.** In addition to the direction of movement the camera also has two viewing angles, one straight down the hallway and the other at a fixed angle pointed towards the left wall. **C.** Schematically we see the direction of viewing for different phases in the stimulus. Note the absolute angle changes are exaggerated. **D.** Different views corresponding to different segments. The straight segment is viewed, as in **B**, directly down the hallway. The turned stimulus shows the view while looking at the left wall with a 15 degree angle. The transition frame shows the view where actively moving from one angle to the other.

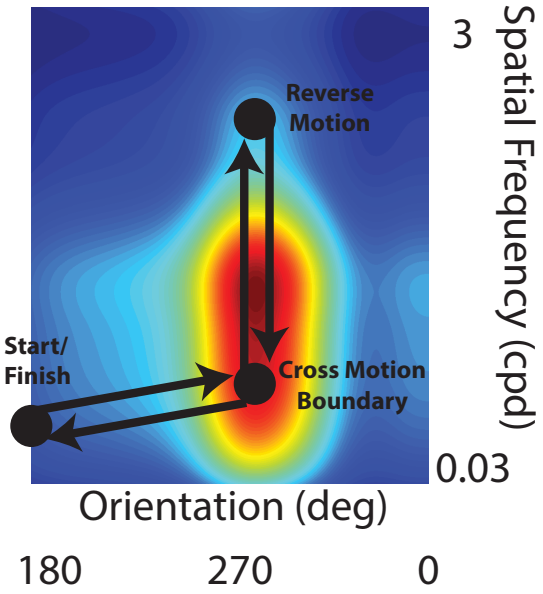

Figure 4.3. Example Parameter Trajectory. From a starting point local spatial frequency and orientation have particular parameters, here projected on top of an example tuning surface to illustrate parameter-driven mean firing rates. As head angle changes during scene exploration both orientation and spatial frequency can cover large areas, resulting in significant changes in mean firing rate.

vastly different firing rates (color of the underlying tuning surface) throughout time.

To perform parameter-driven encoding estimation, we established two similar methods to represent the firing of an individual neuron (see Methods and Figure 4.4). Conceptually, the model involves the assumption that local stimulus properties determine the mean response of the neuron and so accurate tuning surfaces provide insight into how neurons respond to varying stimuli. In a simplified version of the model, if we know where a neuron's receptive field is in visual space (as in Figure 4.4A, left), then we can determine both the stimulus orientation and spatial frequency locally for that receptive field (Figure 4.4A, right). We can easily determine these parameters for every point in space since we have created the stimulus ourselves and defined the parameters (i.e. we have ground truth). As the observer "navigates" the environment, these parameters may be static (as orientation is in this case) or varying (as with spatial frequency), but at any given time t_0 we know exactly what

these parameters are. For each neuron, we also have a tuning surface at a number of different orientations and spatial frequency combinations (shown interpolated in Figure 4.4B). Thus for any particular time t_0, we can use the tuning surface as a "lookup table" to determine what the expected firing rate would be for the observed parameter pairing and generate an expected response at that time $A(t_0)$. Being a sinusoidal grating, the excitation is itself a time-varying value, the temporal aspect of which is reasonably well captured by the local luminance in the receptive field center, shown in Figure 4.4C, top. We use the luminance from the center of the receptive field because this is the only luminance-based measure that captures the modulation caused by the sinusoidal nature of the wall texture without affecting any potential relationship between ON and OFF subfields. As we are using the luminance to provide our temporal modulation, when we capture the value $A(t_0)$ to generate a prediction of average neuron response, we actually capture the peak response to our tuning mapping stimuli rather than the mean response, such that when we perform the final encoding step $L(t_0) * A(t_0)$, the resultant value at any time is meant to capture the observed peri-stimulus time histogram measured response.

The above method illustrates a simplified way to estimate the neural response, but the simplification ignores the fact that spatial parameters vary with space. To accommodate this spatial variance, we modified the encoding method to produce an estimate of the firing rate based on the average parameters across the entire receptive field as shown in Figure 4.5. In this, we can see that an individual receptive field subfield contains numerous individual pixels. Depending on the actual location within the visual stimulus, the orientation and spatial frequency could vary significantly from one end of this subfield to the other or they could not vary at all. Regardless, each pixel is associated with a pair of parameters and we can compute the average parameter values across the entire space before performing our tuning surface lookup to determine $A(t_0)$. Although not shown here, in practice there is little difference

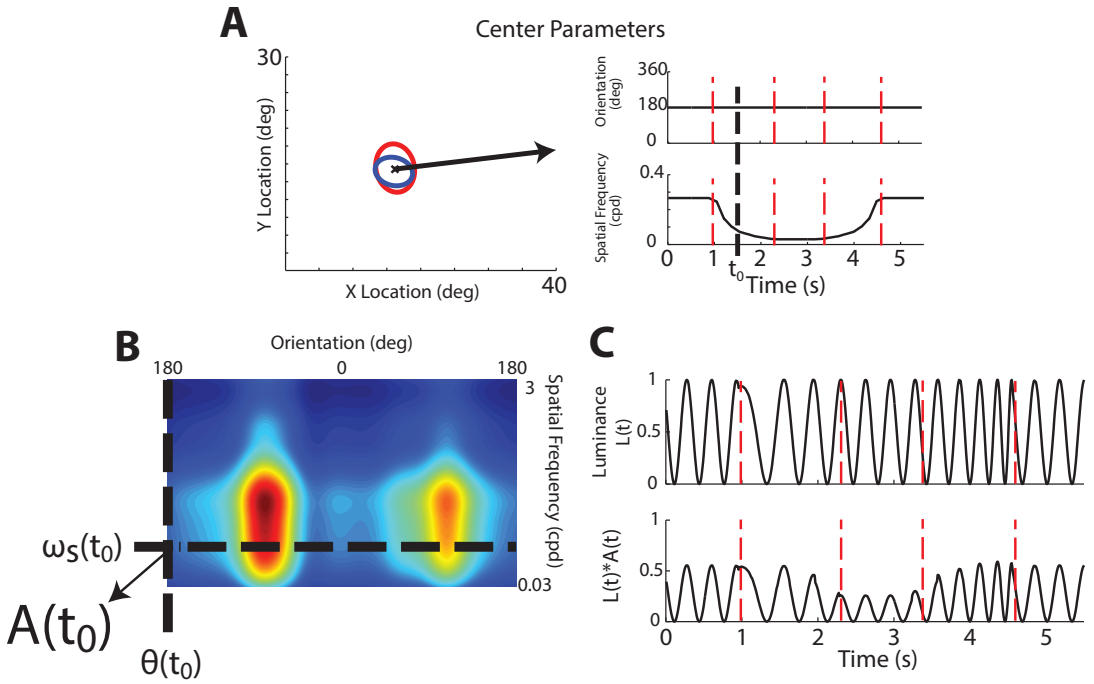

Figure 4.4. Encoding Methodology. A. Encoding is estimated first by identifying the time-varying spatial parameters located at the center of the overall receptive field, here shown in red (ON) and blue (OFF) contours. We see that the stimulus parameters change as a function of time, and identify a particular time t_0 as a "target". **B.** For every stimulus frame, parameter estimates will lead to some unique coordinate on the orientation-spatial frequency tuning surface. This coordinate will tell us the expected neural response associated with this coordinate pair. **C.** Multiplying each of these per-frame estimates by the luminance at the center of the receptive field (top) yields a time-varying estimate of the neural response (bottom). Red dashed lines show the boundaries between the different stimulus segments.

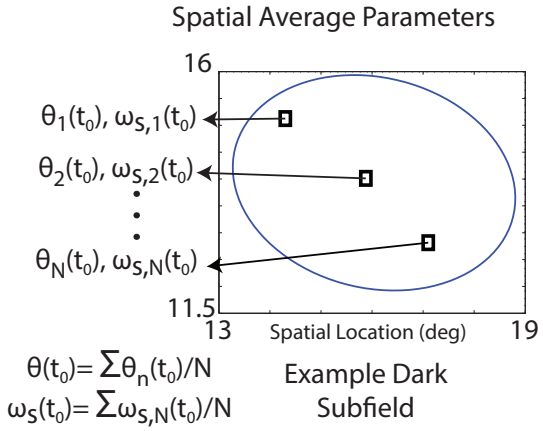

Figure 4.5. Encoding Using Spatial Averages. Each pixel within a receptive field is associated with a potentially unique pairing of spatial frequency and orientation which is known. The average spatial frequency and orientation across all pixels within the receptive field determine what pairing represents the response of the neuron as a whole.

in the accuracy of predictions between this area-based weighting scheme and choosing the parameter values from the center of the receptive field. We could also have used a scheme that estimates the neural response by weighting each individual pixel's predicted response by the strength of the receptive field at that pixel, but due to symmetry of the receptive field size and relatively linear changes in response to parameter changes across the receptive field space, this scheme would produce nearly identical results to using just the center parameters.

The accuracy of this model was quantified via the correlation coefficient between the observed PSTH (binned at the framerate, 120 Hz; 8.33 ms) and this smooth estimate of the neural response. The correlation coefficient is normalized by dividing by the underlying reliability of the recorded PSTH, which was estimated by splitting all recorded trials into two groups and calculating the correlation coefficient between these two groups. This normalization creates a metric similar to percentage of variance explained metrics. Examples of the encoding model results can be seen in Figure 4.6. For each of these encoding estimates, we used the spatially averaged model as described above to predict the neural response. We see accurate encoding predictions

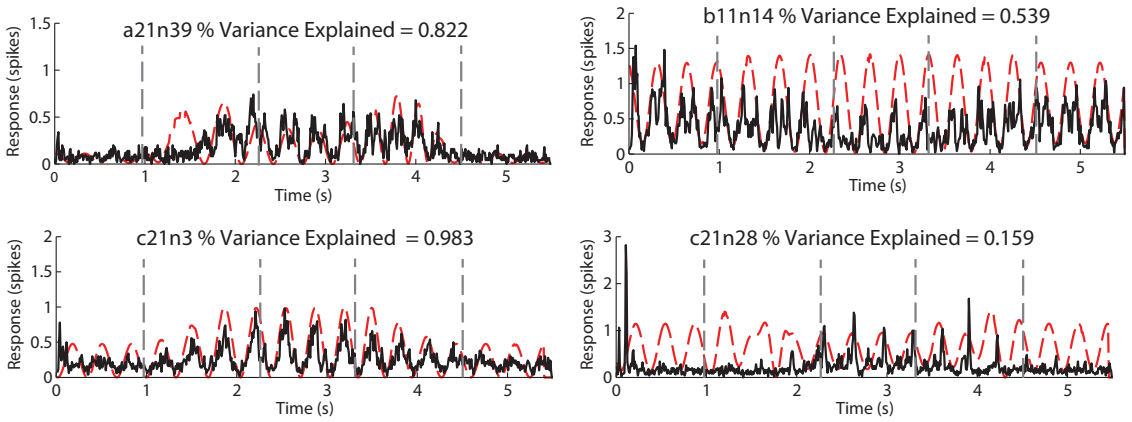

Figure 4.6. Examples of Encoding of Single Neuron Hallway Responses. For each example the prediction coefficient between the observed PSTH (solid black) and the predicted PSTH (dashed red) is shown (see Methods). On the left are two predictions classified as accurate and on the right are two classified as inaccurate. In each case the PSTH reliability was greater than 0.5. Gray bars show separation between stimulus segments (see Figure 4.2).

on the left for two neurons and inaccurate predictions on the right for two neurons. These examples come from multiple simultaneously recorded groupings and all examples have a reliability (see above) greater than 0.5. In the accurate cases it is clear that by and large the encoder prediction matches the timing and relative changes in excitation for these neurons while in the inaccurate cases it is often the result that the prediction fails to encode any change in modulation throughout the stimulus where there actually is significant modulation. The statistics of correlation and reliability are shown in Figure 4.7 for situations where we have removed unreliable units (top) and have not (bottom). In both cases it is clear that most of the predictions are quite poor while the reliability of units uniformly varies from an autocorrelation of 0 to 1.

Given that the location of interesting visual phenomena within the stimulus are fairly clearly defined, we might expect to see spatial patterns to the locations of encoding errors. For example, if local abrupt changes in orientation affected encoding predictions we would expect to see an accumulation of errors near the boundaries where different walls meet. The encoding results are summarized in Figure 4.8A as it

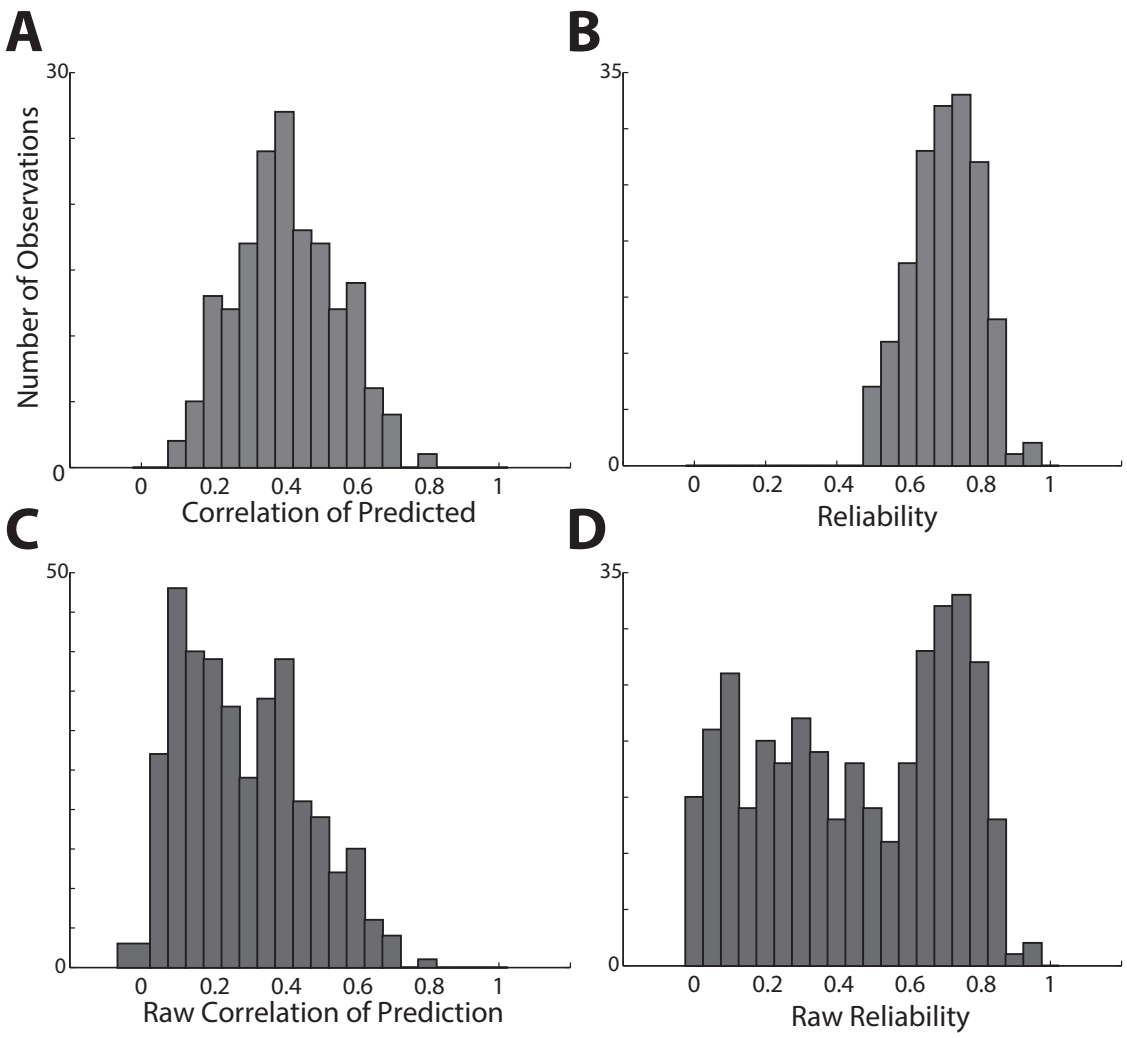

Figure 4.7. Summary of Encoding Performance. **A** After removing unreliable units (see Methods, panel **D**), the correlation between the predicted and observed response is generally quite low, with a clear normal distribution. **B.** After removing unreliable units, the remaining neurons have a roughly uniform distribution of reliability. **C.** Without removing unreliable units we see that in general the correlation between predicted and observed responses is very low. **D.** Including all units tested the distribution of reliability was still fairly uniform.

relates to spatial location in the stimulus. Each dot represents the center of a single neurons receptive field and the color of the dot relates the accuracy of the underlying encoding prediction. Dots are only present for responses in which the reliability was greater than 0.5 in order to remove spuriously poor predictions, leaving 172 of the starting 365 channels (48%) as this analysis used responses from only one eye. Dark blue dots indicate inaccurately predicted encodings while dark red dots indicate perfectly or nearly perfectly predicted encoding. In general the encoding results do not appear to imply that any particular spatial location, and thus natural-like element, is predictive of encoding quality. Across all reliable channels we see that the distribution of the encoding quality is spread across the range of poor accuracy to very high accuracy, shown in Figure 4.8B. The mean prediction coefficient is 0.58 ± 0.19, indicating that the average encoding is a poor representation of the PSTH but with many examples performing both very poorly and very well.

Even though we were unable to distinguish a particular feature or set of features in our natural-like scene that were strongly related to errors in encoding, it remains true that the stimulus is dominated by two particular visual phenomena that are more prevalent in natural scenes. These two features are non-uniform spatial frequency within the image (with at times a very high spatial derivative of spatial frequency) and motion boundaries where two orthogonal drifting orientations meet. It is unclear through existing literature, models, or our observed responses to the hallway stimulus, how these features may challenge common linear tuning-function driven approximations of the neural response.

4.3.3 Decoding of Hallway Parameters

Although single neuron encoding models driven by tuning functions appear to be poor indicators of actual neural responses to our natural-like scene (and perhaps natural scenes in general), it is clear from previous research (Abbott and Dayan,

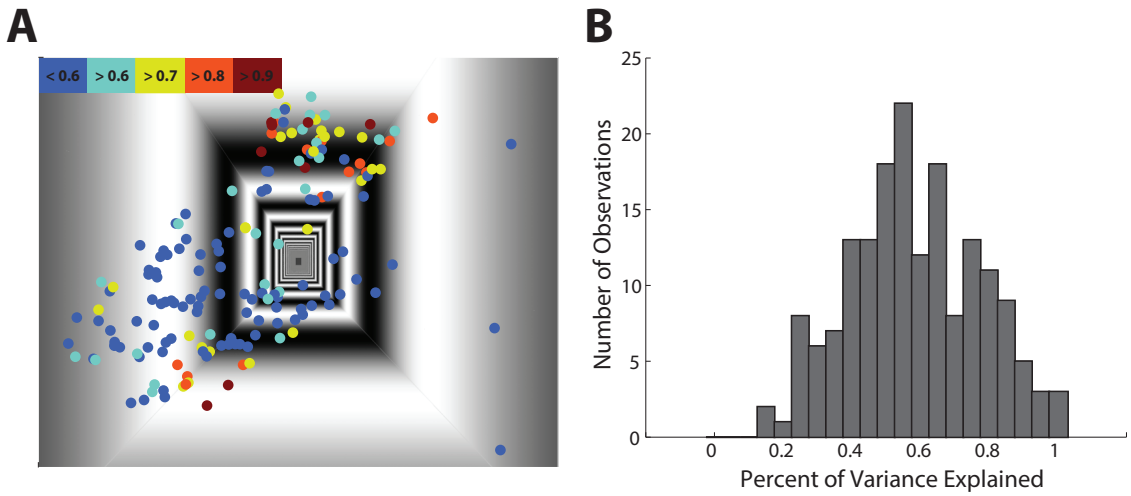

Figure 4.8. Encoding Accuracy as a Function of Spatial Location. A. For all recorded neurons which were also reliable the prediction coefficient is shown as a function of space. Dots are located at the center of a neuron receptive field and the dot color indicates the quality of the response estimate. Dark blue dots indicate a neuron whose response was poorly predicted and maroon dots are neurons which are predicted nearly as well as if the recorded PSTH itself had been used. **B.** The distribution of prediction coefficients shows that the quality of encoding estimates ranges widely, with an average percent of variance explained of 0.58. The distribution contains only units with reliable recordings.

1999, Pouget et al., 1998, Deneve et al., 1999) that populations jointly transmitting information about stimuli are much more likely to be reasonable indicators of stimulus information. Given what we have already found about our natural-like scene, how well can we actually decode what the underlying stimulus parameters are?

We first ask whether or not we can effectively decode the true underlying stimulus parameters that were presented. In a naïve understanding of neural decoding and information transmission, we might expect that higher-level perception of motion within a scene requires an accurate model and representation of parameters like temporal frequency, spatial frequency, and orientation. Accurate decoding of these in primary visual cortex populations, for example, might drive features of area MT motion feature selectivity. Decoding parameterizations of complex natural scenes is likely to be incredibly difficult because determination of the ground truth parameters is a hard task, particularly if the stimulus is arbitrary, like commonly used Hollywood movies or sequences of outdoor and indoor imagery. For our natural-like scene, however, determination of ground truth parameters is relatively straightforward.

4.3.3.1 Decoding Model

We created a parameter decoding model using a population maximum likelihood approach performed for every pixel within the stimulus for each of the two unique segments, straight and turned, illustrated in Figure 4.9. Shown is a simplification of spatial decoding where only two receptive fields overlap, but in practice as many of 80 receptive fields overlapped for at least 1 pixel. For the first receptive field, shown with a dashed ellipse, all pixels within the stimulus attributed to just its receptive field have identical likelihood functions $\mathcal{L}_1 = P(R_1|\theta, \omega_s)$, as a single neuron (and thus RF) can only be associated with a single observed response. For that observed response R_1, we can see that the likelihood function leads to some ambiguity as to which stimulus parameters were presented within those pixel. In the RF of a second

neuron (solid ellipse), the pixels are associated with a similar but different likelihood function $\mathcal{L}_2 = P(R_2|\theta, \omega_s)$ based on the tuning properties of this neuron. Again this likelihood function has some ambiguity associated with it. For visual locations where these two receptive fields overlap, assuming that all channels are conditionally independent, we multiply these two likelihood functions $\mathcal{L}_{RF_1+RF_2} = \mathcal{L}_1 * \mathcal{L}_2$. In this illustration, this has effectively eliminated the ambiguity and given the correct pairing of stimulus parameters that were presented. For the larger population recorded here, the individual likelihood functions are more complex but the effect of aggregating them remains consistently more accurate.

4.3.3.2 Independence of Sampled Populations

Although a range of decoding studies have implicitly made the assumption of population independence, it is important to investigate this assumption directly. This is particularly important here given that we have already shown feature selectivity in cortical neurons is driven by convergent input from a population of thalamic neurons; this arrangement makes it more likely to observe correlations that arise not from common visual space (and thus common stimuli) but from intrinsic correlations based upon common noise. The effect of such correlations is not clear (Abbott and Dayan, 1999, Josić et al., 2009, Cohen and Kohn, 2011, Averbeck and Lee, 2006, Pillow et al., 2008, Klam et al., 2008), but in general research has shown that failing to account for correlations decreases available information and decoding accuracy. Most clearly, Averbeck and Lee (2006) analyzed the effects of accounting for noise correlations on decoding and the effects of ignoring correlations on decoding and found that ignoring such correlations leads to a very small reduction in information if there is any change at all. However, they do show that theoretically a decoder that ignores correlations sets a lower bound on performance. Knowing the magnitude of our pair-wise correlations is thus important. This finding was similar to that found in Johnson and Ray

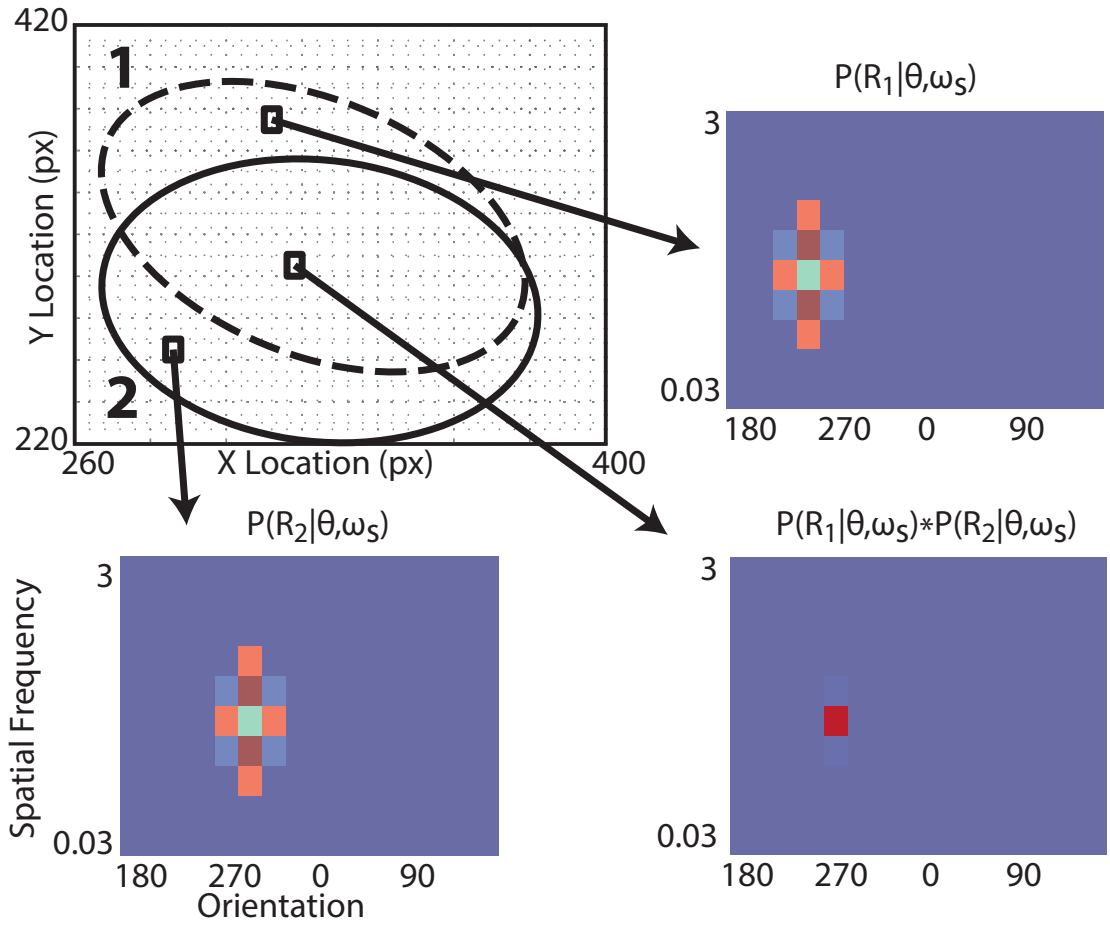

Figure 4.9. Spatial Decoding Method. A. When multiple receptive fields overlap, likelihood matrices interact. In the space where pixels are within only receptive field 1 (dashed line) they are associated with a likelihood $\mathcal{L}_1 = P(R_1|\theta, \omega_s)$. In space where pixels are within only receptive field 2 (solid line), they are they are associated with a likelihood $\mathcal{L}_2 = P(R_2|\theta, \omega_s)$. Pixels which are within both receptive fields have a likelihood equation that the multiplication of both, $\mathcal{L}_{RF_1+RF_2} = \mathcal{L}_1 * \mathcal{L}_2$.

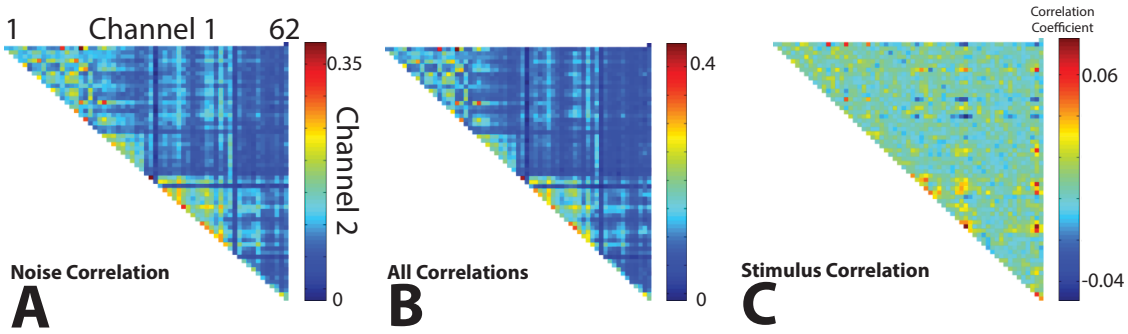

Figure 4.10. Method to Determine Pairwise Correlation. A. Noise correlations between all channels within a single penetration (62 channels). Any correlations caused by the stimulus have been subtracted out. Colormap is scaled independently for each panel to accentuate any patterns that would be lost on a global colormap scaling. **B.** Stimulus-evoked and noise-evoked correlations between all channels. **C.** Stimulus-evoked correlations between all channels. These are segregated by shuffling the order of trials on both neurons when calculating the correlation; this will remove any sources of noise correlation. Note that A = B - C.

(2003) where correlations maximized the Fisher Information in rate codes, but only minimally so compared to having no correlations.

We use standard methods to estimate the noise correlations between each of our channels; in this case "channel" is synonymous with a multiunit recording(see Methods). With this method we found that our recorded neurons show a relatively low level of noise correlation across pairs, as shown in Figure 4.10A. Each individual pixel within the triangle shows the noise correlation $C_{ij,N}$ for a given pair of neurons with a color mapped from dark blue (usually near zero correlation, since no negative correlations were observed) and dark red (highest correlation within a panel). This value is estimated by taking the average single trial pair-wise correlation between two channels, shown in Figure 4.10B and subtracting out the average single trial-shuffled pair-wise correlations, shown in Figure 4.10C.

For this population we see that this procedure shows two things. First, stimulus correlations are almost nonexistent for most possible pairings (1891 possible pairs per simultaneously recorded group). This implies that outside of any possible noise correlations, we have sampled the stimulus space fairly orthogonally in the dimensions

of space and tuning properties. Due to this, the hallway stimulus does not drive any two neurons in identical ways. Secondly, the noise correlations themselves are very low. This implies that the neurons within the population are fairly independent. This observation is shared with a number of different recorded populations, as shown in Figure 4.11. As before, the colormaps are scaled independently to make visualizing emergent patterns easier, and columns and rows where there is no data were the result of trials where no or very little data was recorded. The average correlation coefficient across all pairings shown is 0.06 ± 0.06 (N=23064), indicating that we would expect to observe very little change in information content due to ignoring neuron correlations. Without anatomical mappings of the input structure of each cortical neuron, we can only show correlation measures to estimate the independence of all neurons. In the decoding, however, we have explicitly not accounted for any correlations within the population and this puts a lower bound on the accuracy of our decoding. Given that the observed correlations on average (and at peak) are fairly low, we do not expect that accounting for them would dramatically change the observed results.

4.3.3.3 Decoding Results

In decoding we simultaneously estimate both orientation (θ) and spatial frequency (ω_s) yet we will present the results of decoding each individually for clarity. The results of this decoding approach for the straight segment are shown in Figure 4.12. Although directionality is important, across all of the populations we found moderate to no directional selectivity in the tuning of the neurons, so errors in orientation prediction that are incorrect by 180 degrees are treated as correct estimates (i.e. axial discrimination is most important). This results in 90 degree errors, errors for which the predicted orientation was orthogonal to the true orientation, as the worst possible decoding result. In Figure 4.12A, we see the spatial representation of errors in orientation decoding. As a general principle, multiple observations produced a

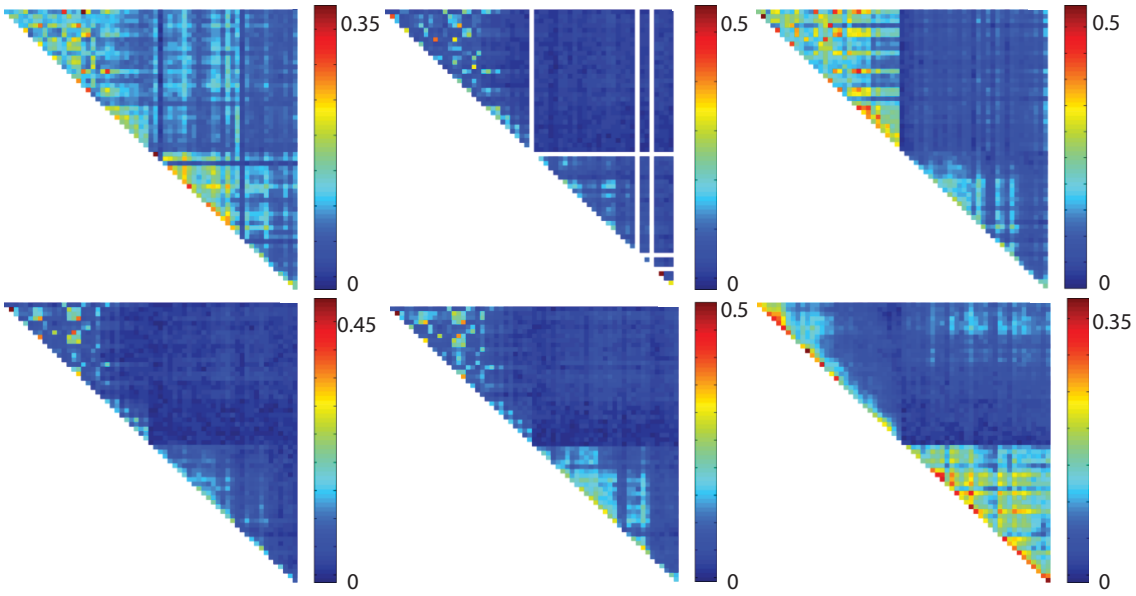

Figure 4.11. Most Recorded Populations Are Highly Uncorrelated. Most populations from our recordings have low noise-induced pairwise correlations. One has significant noise correlations between a number of channels on the first electrode. Here each panel represents a different set of simultaneously recorded channels

more accurate parameter estimate. However, near motion boundaries there are a large amount of orthogonal errors. Figure 4.12B demonstrates that errors are split between 22.5 to 45 degrees incorrect and 45 to 67.5 degrees incorrect with very few orthogonal errors. Spatial frequency estimates, shown in Figure 4.12C, are fairly accurate except in the space surrounding the center of the image where spatial frequencies change most rapidly with space. The distribution of errors decreases roughly exponentially as the true spatial frequency increases with a mean of approximately 0.5 cycles per degree, still representing a significant error. The same analysis for the turned segment was also performed, shown in Figure 4.13. Both spatial frequency estimation error and orientation estimation error are reduced from the straight case.

For spatial locations in which there are multiple receptive fields sampling the stimulus, the decoding results tend to be fairly accurate. This follows a number of previous findings that predict and show the same results (Chen et al., 2006, Quian Quiroga and Panzeri, 2009, Deneve et al., 1999, Johnson and Ray, 2003, Benucci

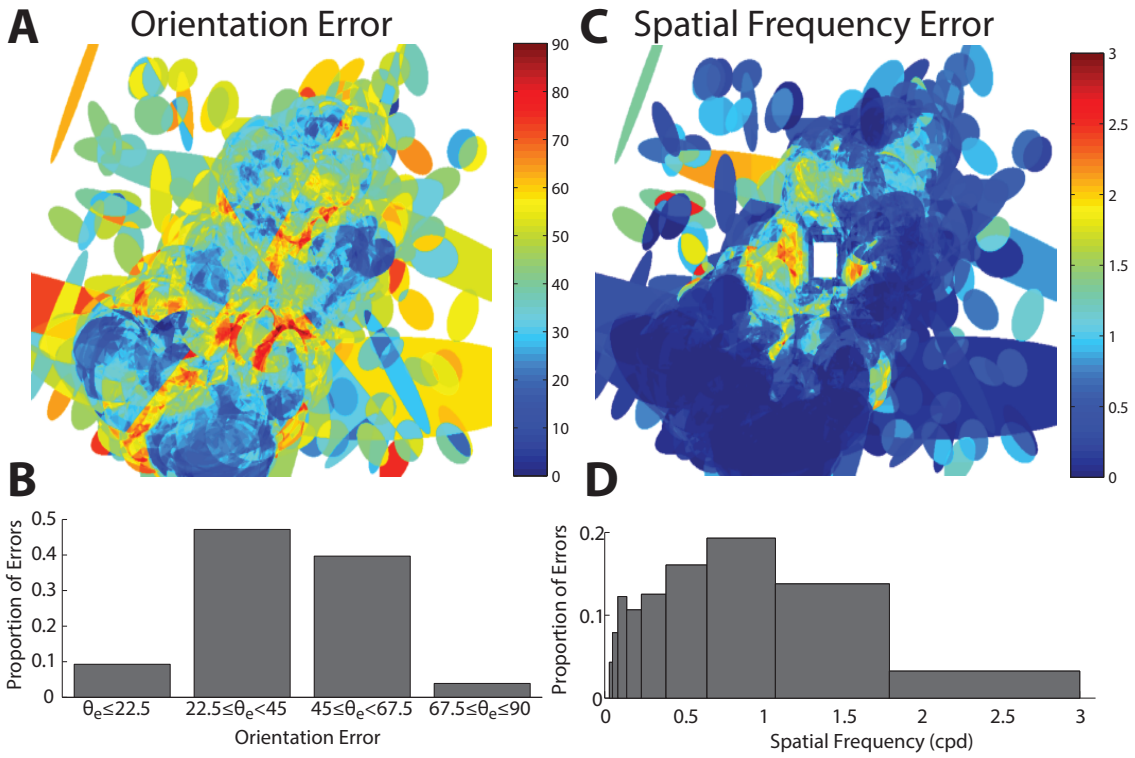

Figure 4.12. Spatial Frequency and Orientation Error for Stimulus Straight Segment. A. Orientation error is mapped to the first quadrant, such that predictions orthogonal to the true axis of movement is consider the worst error type. **B.** Most errors in orientation are less than 45 degrees away from the true direction of motion, but a significant (almost half) are between 45 and 67.5 degrees off. **C.** Spatial frequency errors accumulate near the center of the hallway. **D.** The error has a roughly normal distribution with a mean of approximately 0.52 and a median of 0.35.

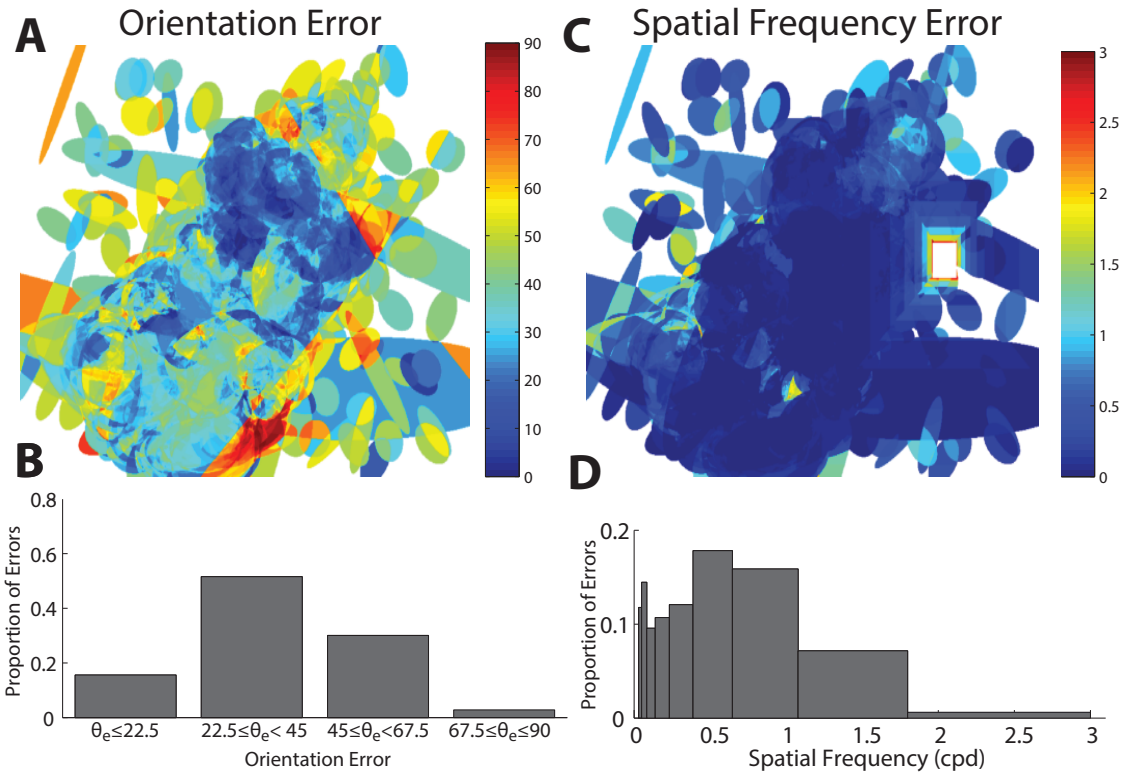

Figure 4.13. Spatial Frequency and Orientation Error for Stimulus Turned Segment. A. Errors are in general lower than when viewing straight. **B.** Most errors are less than 45 degrees. **C.** Spatial frequency error is now significantly lower. **D.** The error distribution has significantly more small errors, with a mean error of 0.32 and a median error of 0.11 cycles per degree.

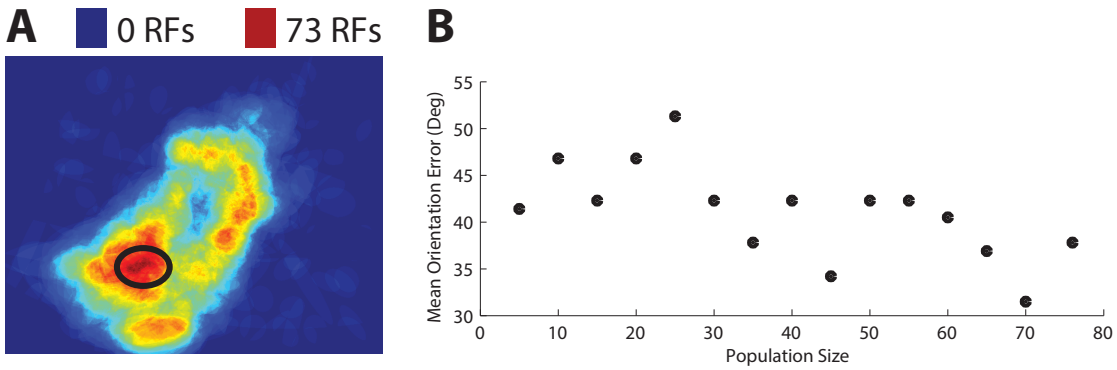

Figure 4.14. Population Size Affects Accuracy of Orientation Decoding. **A.** Heatmap illustrating the number of overlapping receptive fields for each stimulus pixel. The black oval indicates our area of interest. **B.** As the number of neurons within the population is varied, mean error in orientation estimation decreases.

et al., 2009, Abbott and Dayan, 1999) and so from this we can conclude that in general although natural scenes appear to disrupt the encoding of responses in single neurons, this does not strongly appear to negatively affect the ability of the population of such neurons as acting as a reasonably reliable indicator of stimulus properties, provided the population is dense enough. Comparing the results of both Figure 4.12 and Figure 4.13 it is interesting to note that the locations in which qualitatively error is highest for orientation appear around motion boundaries and that locations in which error for spatial frequency are highest are where spatial frequency has a high spatial derivative. To understand how population size affects these predictions, we focus on an area of high overlap (76 overlapping receptive fields) that is positioned near a motion boundary. The location of these receptive fields is indicated in Figure 4.14A, with a black circle indicating the area of interest on top of a heatmap indicating the number of overlapping receptive fields for each pixel. Analyzing just one pixel, we compute the accuracy of decoding stimulus orientation across different compositions of the decoding population, shown in Figure 4.14B. As the number of neurons within the decoding population increases, the mean orientation error decreases.

4.3.4 Segment Classification

It is not necessarily clear, though, that neurons require an accurate determination of exact stimulus properties to maximize information transmission. In addition, the parameter space is more complex than simply predicting stimulus parameters for each pixel; there is an inherent correlation of parameters for nearby locations. Rather than just parameter decoding though, another possibility is that there exists, across neural populations, generalized representations of visual motion caused by exploration and movement in the outside world. While these generalized representations may as a whole capture changes in exact stimulus properties, it is not necessarily to know the underlying properties that created each representation. For a general stimulus, we can consider this to be classification problem: given different segments of motion within a stimulus, how different are the patterns of activation generated within neuron populations across these motions? For the hallway stimulus it is more exact; given a number of possible stimulus choices and a single observed mean firing rate (per neuron), how accurately can we determine which stimulus segment resulted in that observed mean firing rate? The four possible stimulus choices were between forwards and backwards (direction) presentations and straight or turned segments (angle). The framework we used for classification was a maximum likelihood estimator (see Methods), model distributions for which were trained using half of recorded trials and tested using the other half.

Our classification task, shown in Figure 4.15, had two modes. The first restricted available choices to two: either direction or angle was fixed and the classifier had to choose from the two remaining options. In the two example neurons in Figure 4.15A, it is clear that these neurons are sensitive to spatial frequency changes induced by changes in the head angle, just by comparing their raster plots. This observation is partially borne out in Figure 4.15B, where one performs the above two-choice classification task correctly in approximately 80% of trials regardless of which parameter is

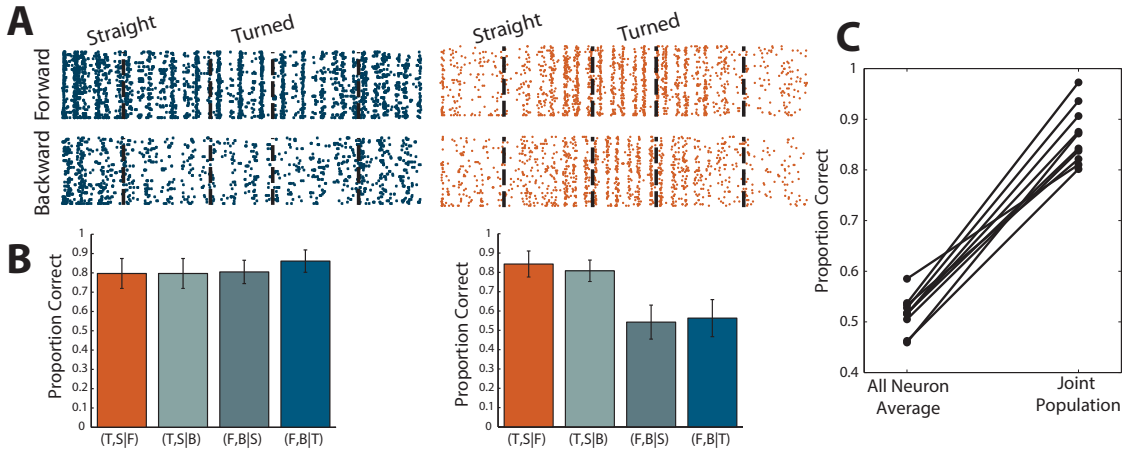

Figure 4.15. Classification Decoding of Single Neurons and Populations. **A.** For two individual neurons the rastergrams for responses to the forward and backwards presentations of the hallway stimulus are shown. Dashed lines indicate boundaries between unique segments within the stimulus. **B.** For a decoding task that tasked an ideal observer to decode between one of two choices given that either translation direction or stimulus segment were fixed the results are shown for the two neurons in A. **C.** The average results across all neurons for a similar classification with four possible choices are shown next to the results from using the joint activity across all neurons.

fixed. The other neuron, however, matches our expectation and classifies head angle with a similar performance but classifies directionality very close to chance (50%). We then make the task harder by removing any restrictions on directionality and angle, restoring the four choices, and pose the same classification problem to both the individual neurons as well as to populations of simultaneously recorded neurons again assuming conditional independence. Estimates were combined jointly across neurons by adding their log likelihood functions and then minimizing to achieve a more accurate estimate. From Figure 4.15C it is clear that while the average neuron performs at a little over twice chance (25%), the joint classification from each population is significantly better, with a minimum performance of 80% and some close to 95% accuracy.

The improvement in decoding and classification accuracy is tied to the number of independent observers of the stimuli. As can be seen in Figure 4.16, across all sets

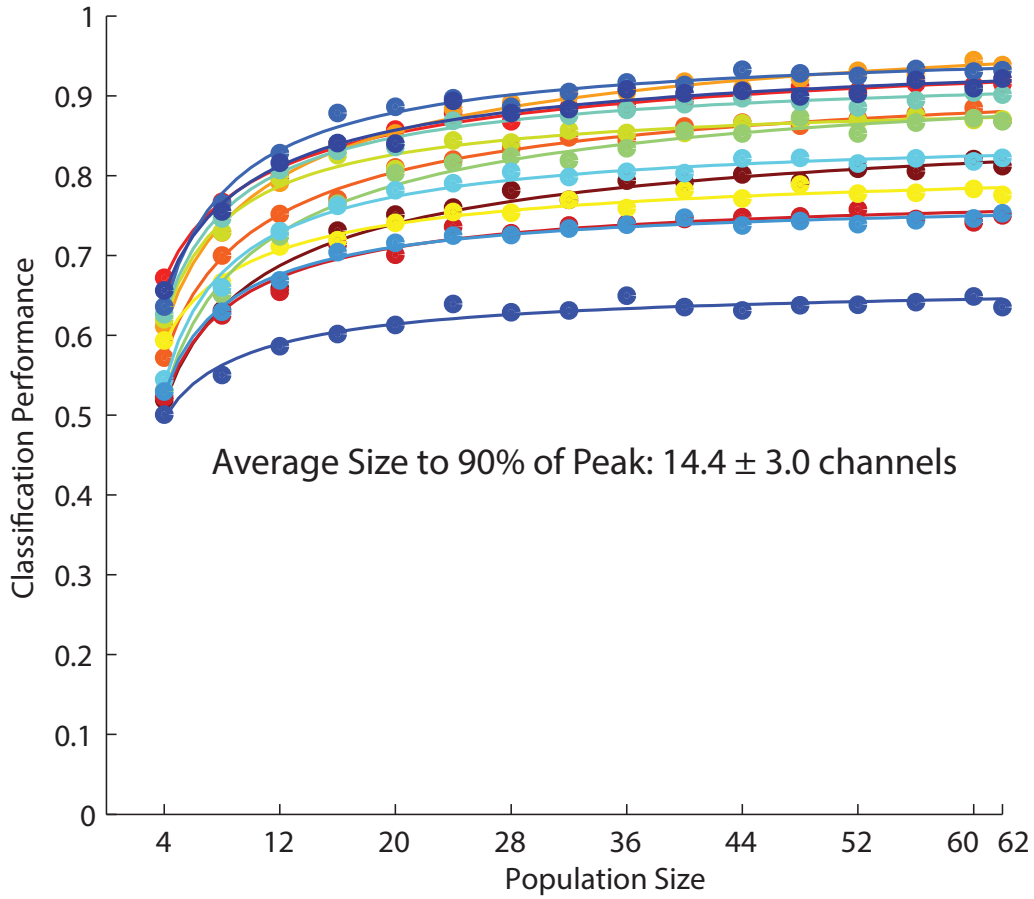

Figure 4.16. Classification Accuracy is Tied to Population Size. Across different simultaneously recorded populations (shown with different color lines), classification accuracy increases in a power-law fashion with increasing population size. Each point represents the average accuracy for each population size and the solid line is a power-law fit.

of simultaneously recorded neurons classification accuracy increases with the number of channels included in the joint calculation, reaching 90% of the full performance with an average of 14.4 ± 3.0 channels, indicating that relatively small numbers of convergent inputs would be required to perform the decoding task with high accuracy.

4.4 Discussion

4.4.1 Summary

We used a novel and self-generated visual stimulus to probe the effects of a "minimally natural" scene on feature selectivity in the cat visual system. This scene simulates a hallway of infinite length with sinusoidal textures mapped to the walls. The properties of the stimulus were designed to match the expected range of tuning properties of cortical V1 neurons. When we used pairings of simultaneously varying orientation and spatial frequency to map the tuning surface, predictions of the neural response based on that tuning surface had a wide distribution of qualities from poor estimates to very accurate estimates. Using these tuning surfaces to decode the underlying stimulus parameters showed some location-based patterns in accuracy across the entire population of recorded channels and classifying different types of motion within the hallway stimulus using the entire population was almost always accurate greater than 80% of the time. These results demonstrated that motion information is present in the population of responses but single neurons have unique and often unpredictable responses to even this "nearly natural" scene.

4.4.2 Stereopsis and The Sinusoidal Hallway

Stereopsis - the perception of depth from two offset two-dimensional sensors - has a history of being analyzed using a variety of sinusoidal gratings, sometimes applying a tilt (Tyler and Sutter, 1979). In general, stereopsis is concerned with methods of binocular disparity where each eye provides subtly different information about orientation or spatial frequency and neurons within the primary visual cortex (and elsewhere) are sensitive to these discrepancies (Siddiqui and Bhaumik, 2013, Ban et al., 2012, Bridge and Cumming, 2001, Liu et al., 2004, Nadler et al., 2013, Sanada and Ohzawa, 2006). In this study we have covered one eye at a time such that the stimulus is always presented monocularly and thus offers no binocular disparity

information. Despite this, we appear to exhibit the worst predictions of motion information in the stimulus in places where these rivalries might be highest; at motion boundaries and spaces of variable spatial frequency. It is possible that the errors we show in decoding and encoding could be attributed to a failure to account for binocular input on some of the recorded neurons. This seems to be an unlikely result given that two experimental steps were taken to remove any binocularity effects. First, as mentioned, one eye was always covered and prevented from receiving visual input. Thus, its contribution to any visually-guided information should be minimal. Secondly for each recording session from each eye, tuning properties were captured such that the tuning results were independent for each eye. Any binocular effects (caused by removal of any visually guided input from one eye) would thus have already been accounted for within the mapping properties. Due to these controls, it appears the errors introduced by the natural perspective are not simply due to failures in accounting for stereopsis.

4.4.3 Populations and Natural Scenes

Our results tell us something very interesting about the practicality of the usage of natural scenes in visual system analysis. The combination of artificial stimuli to measure tuning functions along with the usage of natural scene imagery poses a clear challenge to using these tuning functions to later accurately determine how these neurons code for particular stimulus parameters. In many cases the errors that occur from using this method, even when it accumulates information from the entire receptive field space, are significant. Placing the analysis of natural scenes in a more behaviorally and ethologically relevant context by examining the responses across entire populations, on the other hand, changes the observations considerably. In this case errors are minimized where space is sampled by many different neurons and in classification tasks performance of such neurons is almost perfect. Thus it would

appear as though natural scenes are best employed in scenarios where the hypotheses and questions are geared towards transmission of information in populations rather than single neurons, as the underlying model of neural response in populations appears to be somewhat less sensitive. This is beneficial because tuning function driven models can be simpler and more computationally tractable than receptive field based models under many conditions and may also scale up more elegantly across increasingly large populations. On the other hand, it implies that for single neurons it is necessary to construct a more complex set of tuning functions that actually account for the visual phenomena seen in natural scenes.

CHAPTER V

EXPLORING SPECIFIC NATURAL-LIKE FEATURES

5.1 Introduction

5.1.1 Feature Selectivity and Spatial Change

In the quest to understand how visual neurons respond to the outside world, we commonly describe them in terms of tuning features, as has been previously discussed. We might consider a neuron to be primarily driven by some combination of spatial frequency, temporal frequency, and stimulus orientation to the point that knowledge of these features would give us a highly accurate description of the neural response (Movshon et al., 1978, De Valois et al., 1982a, Priebe et al., 2006). How accurate is that perspective when stimuli are generalized? Do spatially fixed parameters actually describe the complexity of neural responses to stimuli or are they an excessively reduced representation of space? Historically, some permutation of examples have been explored; different spatial and temporal frequency combinations, sometimes one or the other has been fixed, combinations of orientation and spatial frequency and/or temporal frequency, and even up to different phases between binocular presentations of such stimuli (Adelson and Bergen, 1985, Field, 1987, Fleet et al., 1996). In all cases these combinatorial interactions have revealed important understanding of visual coding, but it is not clear how comprehensive this understanding is.

It is clear that natural scenes and similar constructs introduce visual elements that can challenge how well under-specified feature selectivity properties capture neural and population responses. In Chapter 3, I showed that even for a very simple version of a natural scene we had difficulty representing the neural response. These challenges

translate into reduced property decoding performance but not reduced motion classification performance. What is it about these scenes, though, that actually causes this to happen? While Figure 4.8 fails to provide any insight into important features, an understanding of the natural world and effects like perspective and object motion provide clues to the kinds of features that must be investigated.

5.1.2 Two Natural Features

More than just being the features that are present in the hallway stimulus, perspective and motion boundaries are also incredibly important to natural vision in general. In complex stimuli these features get intricately involved with each other, as physical properties of objects and their relationships to each other get transformed by the effect of perspective. Consider these features in isolation as shown in Figure 5.1. Spatial frequency changes (as in Figure 5.1A) that occur within local space are a key aspect of perspective, one of the fundamental building blocks of the outside world. Even if each patch had the same central spatial frequency, the fact that it changes so significantly across the given view of space (roughly equal to the size of a V1 receptive field) would appear to have interesting implications for the neural response. Even when looking through a forest, trees which are far away will have higher spatial frequencies and converge to a horizon. The same is true when looking down a street in a large city, increasing spatial frequency and convergence to a horizon. Perspective is so fundamental to natural vision that it is crucial to understand how it induced unique responses in single neurons in visual cortex.

Motion boundaries on the other hand describe the way things actually move in the outside world. As we can see in Figure 5.1B, even the act of rotating a camera can explore a boundary between two unique motion fields (assume each field was animated and drifting) in a useful way. Rather than sets of coherent or incoherent dots or plaids as we may find in common MT and MST literature, natural scenes have

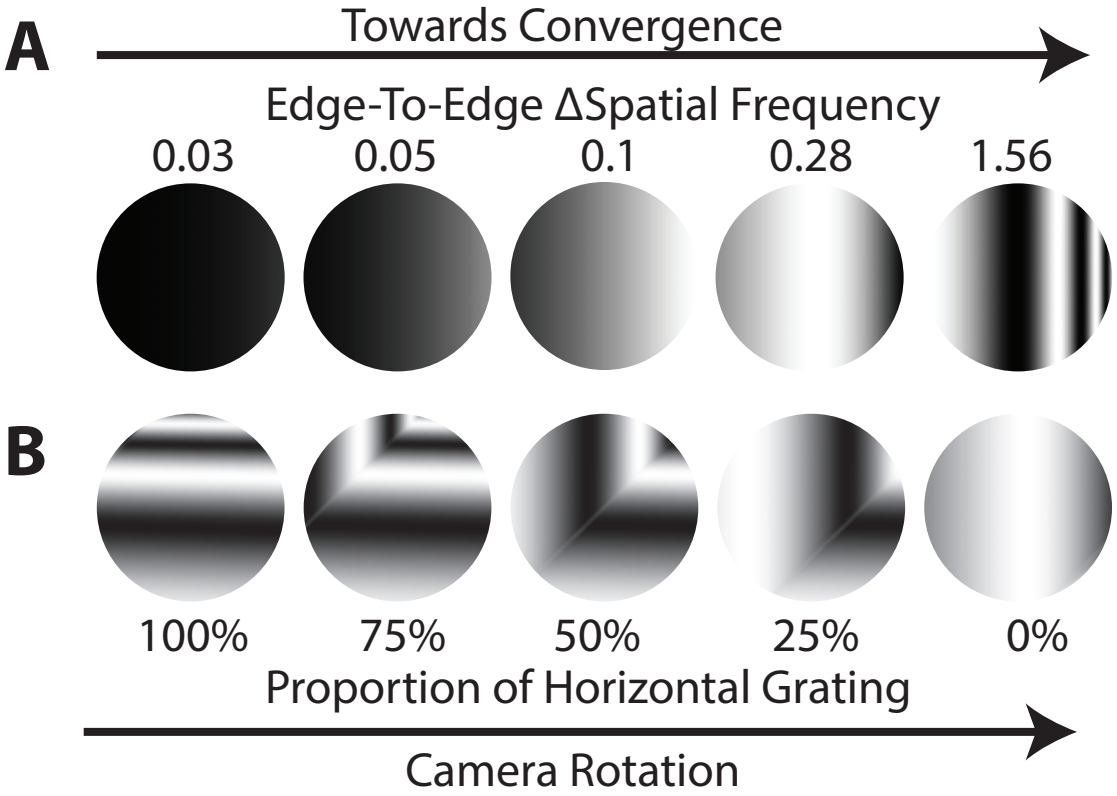

Figure 5.1. Two Unique and Visually Important Features. A. Within the space of a primary visual cortex receptive field perspective alters the local spatial frequency changes based on the distance from the convergence point (center of gaze). Locations closer to convergence have a significantly higher change in spatial frequency from one edge of the window to the other. **B.** Motion boundaries alter the local orientation (and in more complex scenarios spatial frequency) content between two or more unique motion fields. Even self-motion within the environment can cause a shift in the location of these boundaries.

objects that move against each other, objects which move against backgrounds, and even objects that move against each other with transparency. Motion boundaries have been found to be omnipresent in natural scenes (Roth and Black, 2007) and have been observed to drive responses in multiple visual cortical areas (Larsson et al., 2010). Just as for tilt angles, understanding how these motion boundaries, even in simplistic cases like our hallway, mediate the response between the two unique fields is important for understanding information transmission about motion.

These features also have unique applications in visual neurons as they are commonly applied to the analysis of binocular information within scenes. Some neurons in primary visual cortex receive information from both eyes at once and studies commonly show that this information leads to tuning which is selective for orientation or spatial frequency differences between the eyes (Siddiqui and Bhaumik, 2013, Ban et al., 2012, Bridge and Cumming, 2001, Liu et al., 2004, Nadler et al., 2013, Sanada and Ohzawa, 2006). These differences correspond to the features we have embedded in the hallway stimulus, with orientation differences indicating motion boundaries and spatial frequency differences indicating planar tilt (or perspective). Since these features are almost only ever applied to binocular information, what are the implications for feature selectivity to these natural features under conditions of monocular stimulation?

To explore these two unique features, we created a set of new simple stimuli that simulate the natural features singularly. The first is created through drifting gratings that tilt within the screen, changing the spatial frequency from one end of the image to another. The second is orthogonally oriented drifting gratings with a boundary between the two that moves in visual space. Analyzing the responses to each of these stimuli we find that both the tilt stimuli and the motion boundary reveal statistically significant changes in response from classical predictions based on feature selectivity. Following these significant changes, we compared decoding of each type of stimuli to verify that the observations truly conveyed additional information and found that decoding performance was significantly improved compared to what we would expect from tuning properties. These results together indicate that responses to specific features within natural scenes contain more information about the stimuli than typical aspects of neural feature selectivity would predict. This changes the perspective on how we simplify visual neurons, suggesting that future models of these neurons requires a more in-depth representation of feature selectivity to include the effects

of natural elements like perspective and motion fields. Interestingly these additional tuning properties appear to overlap with representations of neural responses typically reserved for binocular information.

5.2 Methods

5.2.1 Electrophysiology

Neural recordings were captured from primary visual cortex of anesthetized cats in the same manner as in Chapter 3. Electrophysiological recordings were made using dual Neuronexus 32-electrode probes in one cat, one placed perpendicular to the cortex and a second placed approximately parallel. The perpendicular probe was aimed to traverse down an orientation column while the parallel probe was aimed to cross multiple columns. Each electrode site was separated by 100 um for a total length of 3.2 mm with a total of 384 recorded channels across all penetrations. In another cat three 32-channel probes were placed, with two in one hemisphere and the third in the other hemisphere, with 288 total recorded channels.

5.2.2 Visual Stimuli

Visual stimuli were presented on a CRT monitor at 120 Hz with presented images gamma corrected to ensure a linear representation of the entire white to black range of the monitor. Spatial frequency and orientation tuning properties were simultaneously measured using a series of drifting sinusoidal gratings that swept across a parameter space of both spatial frequency and orientation. Orientation was limited to 0, 90, 180, and 270 degrees with the spatial frequency ranging from 0.01 to 1 or 0.03 to 1.5 cycles per degree (with 10 logarithmically spaced increments) with 60 total trials for each orientation-spatial frequency pair. The spatial resolution for all stimuli was 0.056 degrees per pixel. The temporal frequency for all drifting gratings was 5 Hz. All stimuli were presented at 100% contrast.

The planar tilt stimuli were presented in each of 4 cardinal orientations with 9

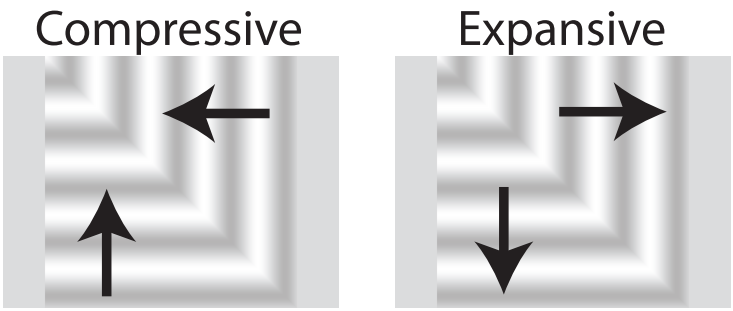

Figure 5.2. Compressive and Expansive Motion Boundary Stimuli. Compressive motion boundary stimuli have motion directions that point towards their shared motion boundary while expansive stimuli have motion directions that point away from their shared motion boundary.

possible choices from both positive and negative tilt angles to capture both increases and decreases in local spatial frequency. The motion boundary stimulus had 19 possible stimuli each placing the boundary between orthogonally oriented motion fields of identical spatial frequency in a different location, moving from bottom-left to top-right. The motion boundary stimuli are presented as compressive (drifting gratings move towards the boundary between them) or expansive (drifting gratings move away from the boundary between them). To understand the difference between these two, reference the illustration in Figure 5.2. For the planar tilt and motion boundary stimulus presentations each of the possible choices were presented in a fixed random order and preceded by 200 ms of blank visual stimuli (50% gray). Both stimuli were presented at 120 Hz and 100% contrast.

5.2.3 Data Analysis

Electrophysiological data was captured for two different animals and identical analysis was performed on both sets of data.

5.2.3.1 Receptive Fields

Individual ON and OFF subfields were mapped using sequences of dark and light sparse noise stimuli (Jones and Palmer, 1987) with approximately 5 minutes of stimuli.

Using a standard reverse correlation technique the recorded spike times were used to generate the receptive field kernels. The kernels were mapped to a 50% (half height) contour and then fit with an ellipse to obtain a smooth representation of the receptive field. The size of each RF was calculated as the average radius between the ON and OFF subfields; in nearly every case the ON and OFF subfields were located on top of each other so the average radius captures the effective receptive field size. The mean receptive field radius was 1.93 ± 0.60 degrees.

5.2.3.2 Spatial Frequency Tuning

Spatial frequency tuning was fit using a spline approximation using the fit function in MATLAB. This captured a smooth and accurate approximation to sub-sample the spatial frequency space at a granularity finer than we were able to record. To approximate the statistics of the spatial frequency tuning, each tuning curve was fit with a Gaussian, providing a fairly accurate fit. The average center frequency was 0.84 ± 0.53 cycles per degree and the average tuning width was 0.87 ± 0.62 cycles per degree.

5.2.3.3 Directional Index

Directional Index was also computed for each channel (as in Stanley et al. (2012)), using the equation

$$DI = 1 - \frac{R_{\theta_0+180}}{R_{\theta_0}}$$

where θ_0 is the preferred orientation. A directional index value of 0 indicates that there is no directional selectivity and axially identical stimuli (for example vertical gratings that drift leftward or rightward) would elicit identical responses. A directional index value of 1 indicates that the response is perfectly directionally selective. The mean directional index for the recordings shown was 0.39 ± 0.28, indicating as a whole the recordings are poorly directionally selective.

5.2.3.4 Tilt Stimulus

The response to each stimulus was classified by calculating the mean response to each tilt angle. In order to measure the actual effect of each tilt angle the results are displayed as a "delta from untilted" response, where ϕ is the angle of tilt,

$$\Delta R_\phi = |R_\phi - R_0|$$

since the response when the stimulus is un-tilted, R_0 corresponds to a known quantity from the tuning surface measurements. We represent ΔR_ϕ as an absolute value to emphasize that this is a change from un-tilted response calculation.

When comparing the results of the tilt stimulus across multiple different populations (Figure 5.8), comparisons were made between the maximum observed delta values, $\max \Delta R_\phi$, for both the recorded and expected results. In practice this value almost always occurred at $\phi = -45$ deg.

Local spatial frequency for every pixel within each tilt stimuli was calculated numerically, identical to the procedure for the sinusoidal hallway (see Chapter 3 Methods). The results, shown in Figure 5.3, were smooth and continuous, with higher edge-to-edge changes in spatial frequency when the tilt angle was higher. Note that the spatial frequency also changes nonlinearly with half of the image having slowly varying if not invariant local spatial frequency and the other half having large changes.

5.2.3.5 Motion Boundary Stimulus

The response to the motion boundary stimulus is represented by the mean firing rate in response to each boundary location. As the observed response and any trends in firing rate changes with motion boundary location are directly related to the location of the receptive field, we normalized the location of the motion boundary relative to perpendicular distance l and the receptive field radius r. Thus the motion boundary

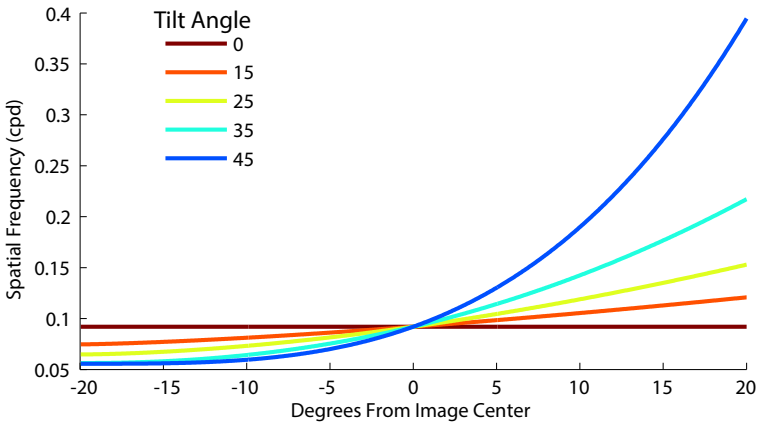

Figure 5.3. Calculated Local Spatial Frequency. For different tilt angles local spatial frequency at every pixel location varies considerably based on the underlying angle. At the center of vision all tilts produce identical spatial frequencies, since the grating is rotated around the center of the image.

location

$$L = \frac{l}{r}$$

and when $L > 1$ the boundary is outside the classical receptive field (measured here at the 50% contour) and when $L < 1$ the boundary is within the classical receptive field.

To compare the error between the observed response and the expected response across all channels, motion boundary location was split into bins of 0.5 radii. The mean error ($R_{\phi,obs} - R_{\phi,expect}$) for all trials in each bin was calculated and significant differences within each bin were determined by comparing the distributions of observed responses and expected responses within each bin.

5.2.4 Tilt Response Prediction

Responses to each different tilt angle were estimated by integration of the parameters across the entire receptive field area. If portions of the receptive field contain spatial frequencies that are significantly higher (or lower) than the spatial frequency at the center, then such a spatial averaging is necessary to account for the true average expected response. Consider for example a neuron which is tuned to high spatial

frequencies and its receptive field has 75% of its area stimulated by a low, fixed spatial frequency and the rest by a high spatial frequency. From the properties in the receptive field center, we might assume the neural response to be low, from the average properties across the receptive field space we would realize it might actually be significantly higher than previously expected.

If the receptive field region is given by S then every element within that space is given by $S(x, y)$. Correspondingly, because the stimulus has been made precisely we can determine stimulus parameters at every pixel within that receptive field as well,

$$S_\theta(x, y), S_{\omega_s}(x, y)$$

We can use this spatial description of the stimulus parameters to generate a prediction of the mean neural firing rate in response to the tilt stimulus by using the tuning surface A across all N elements within the receptive field area

$$\hat{R} = \frac{\sum_{x,y}^{N} A(\theta(x, y), \omega_s(x, y))}{N}$$

There are also alternative approaches to calculating this estimated response $\hat{R}$. One approach would be to additionally weight the spatial integration by the strength of the receptive field at each (x, y) location. Another approach would be simply to use the stimulus parameters from the center of the receptive field. If the receptive field is sufficiently peaked then these approaches are identical. We chose to implement the spatial average technique to capture the greatest possible variance with space, but in practice each of the approaches was almost identical given that the receptive fields are roughly symmetrical and the tuning surfaces are monotonic and fairly linear over the appropriate range.

5.2.5 Motion Boundary Prediction

The response to different motion boundary locations was predicted on the basis of integration of orientation within the classical receptive field. Note that spatial frequency is constant across the entire stimulus space. The stimulus is composed of two discrete orientations, meaning there is a fraction of receptive field area in one field and the complementary fraction of area in the other field. If we keep our notation as above that S is the receptive field region, then $S = S_1 + S_2$ where S_1 describes the receptive field region in motion field 1 and the same for S_2, such that

$$\alpha_{\theta_1} = \frac{S_1}{S}$$

$$\alpha_{\theta_2} = \frac{S_2}{S}$$

$$\alpha_{\theta_1} + \alpha_{\theta_2} = 1$$

The expected response for each motion boundary location was calculated by associating these fractions with the response elicited by the spatial frequency and orientation parameters from the tuning surface

$$\hat{R_L} = \alpha_{\theta_1}\hat{R}(\theta_1, \omega_s) + \alpha_{\theta_2}\hat{R}(\theta_2, \omega_s)$$

However a first analysis revealed that additional error was introduced due to $\hat{R}(\theta, \omega_s)$ being generally poor predictors of the response when $\alpha = 1$ as there was typically an uncorrected offset between the observed mean response and the expected mean response. To correct for this the expected response $\hat{R_L}$ was forced to be equal to the true observed response R_L whenever the stimulus was a full-frame single grating ($\alpha = 1$).

5.2.6 Tilt Stimulus Classification

Random single trial mean responses were classified as having resulted from one of the nine possible tilt angles presented for a single orientation of stimulus. From half

of the recorded trials for each possible tilt angle an empirical response distribution ($D(\phi)$) was calculated. This distribution was not fit with any descriptive function. From the remaining half of trials a single response (r_n) was randomly chosen. The likelihood of a particular tilt angle having resulted in the observed response was given by

$$\mathcal{L} = P(r_n|\phi)$$

A single channel generates a predicted tilt angle by finding the tilt angle from all possible values that maximizes this quantity,

$$\widehat{\phi} = \underset{\phi}{\operatorname{argmax}}\, P(r_n|\phi)$$

For computing the predicted tilt angle across multiple independent observations (in this case different channels), the log-likelihood functions for all observations are added.

$$\widehat{\phi} = \underset{\phi}{\operatorname{argmax}} \sum_{n}^{N} \log P(r_n|\phi)$$

These results were compared to a classifier based on the tuning properties of each channel, with the response to the tilt stimuli being extrapolated from the local spatial frequency for each tilt stimulus. As local spatial frequency varied on a continuum finer than was captured from the tuning surface measurements, spline curves were fit to both the mean and the standard deviation at each measured spatial frequency. For each tilt angle the distribution for a single channel was estimated with a Gaussian,

$$\mathcal{N}(\mu(\omega_s), \sigma(\omega_s))$$

Single trials were randomly drawn from this distribution for decoding. Likelihood functions and population estimates were calculated as previously shown.

5.2.7 Motion Boundary Classification

To test classification of motion boundary location we limited the possible classes to 5 from the total of 19 motion boundary stimuli presented. These 5 classes corresponded

to both full-frame drifting gratings (horizontal and vertical) and locations where the boundary was in the bottom left (25%) directly down the middle (50%) and at the top right (75%). The choice was made to limit possible classifications due to relatively small observed changes in most of the boundary locations that were reasonably far away from the classical receptive field. The chosen categories represent spaces far from the receptive field, within the "non-classical" receptive field, and within the classical receptive field for most of the recorded channels. Different choices of which 5 boundary locations to classify would have likely produced different results. The example 5 locations uniformly span the entire set of locations. Distributions and likelihood functions were calculated identically as for the tilt stimuli.

Classification was also performed based on the theoretical performance of a classifier based on tuning properties. The classifier was initially trained by finding the response distribution attributable to both motion fields individually and then re-calculating those distributions based on the the percentage of stimulus pixels from each motion that were present within the receptive field. Each distribution was assumed to be normal

$$\mathcal{N}(\mu(\theta_1), \sigma(\theta_1))$$

For example, when the motion boundary is in the middle of the image (the 50% stimulus) a particular neuron has 80% of its space ($\alpha_{\theta_1} = 0.8$) in a 90 degree orientation drifting grating and the remaining 20% ($\alpha_{\theta_2} = 0.2$) is in a 0 degree orientation drifting grating. We re-calculated the expected response distribution by first drawing a random number (x) between 0 and 1; if this number is less than 0.8 for this example we draw a single observation from the distribution associated with the 90 degree orientation grating (see definitions for α above).

$$r = \begin{cases} \mathcal{N}(\mu(\theta_1), \sigma(\theta_1)) & \text{if } x \leq \alpha_1 \\ \mathcal{N}(\mu(\theta_2), \sigma(\theta_2)) & \text{if } x > \alpha_2 \end{cases}$$

By repeating this process a number of times for every channel over every stimulus (of

the 5 chosen to classify), we can generate new response distributions that include the modulatory activity of the motion boundary. Test single trials are then drawn from a normal distribution (which was empirically verified to be the best-fitting descriptive distribution) with a mean and standard deviation derived from these iteratively determined distributions. The decoding proceeded as above for the maximum likelihood process.

5.3 Results

To address the effect of isolated natural-like features on motion and parameter sensitivity, we created and presented two novel stimuli. These stimuli follow the lead of the sinusoidal hallway and are generated to be specific instantiations of natural-like features. Where even the sinusoidal hallway has some complex amalgam of different visual phenomena, these stimuli are distillations and iterations of a single basic feature.

5.3.1 Response Statistics

The basic response statistics of all units recorded during one experiment are shown in Figure 5.4. As a reduced representation of orientation was mapped using drifting sinusoidal gratings, we could not reliably calculate preferred orientations or orientation tuning width for this population. From Figure 5.4A we can see that receptive fields tended to cluster in a reduced area of the visual space with a mean receptive field radius of 1.93 $\pm$ 0.60 degrees (N = 384). The average center spatial frequency was 0.84 $\pm$ 0.53 cycles per degree and the average tuning width was 0.87 $\pm$ 0.62 cycles per degree. The mean directional index for the recordings shown was 0.39 $\pm$ 0.28, indicating as a whole the population is generally directionally insensitive.

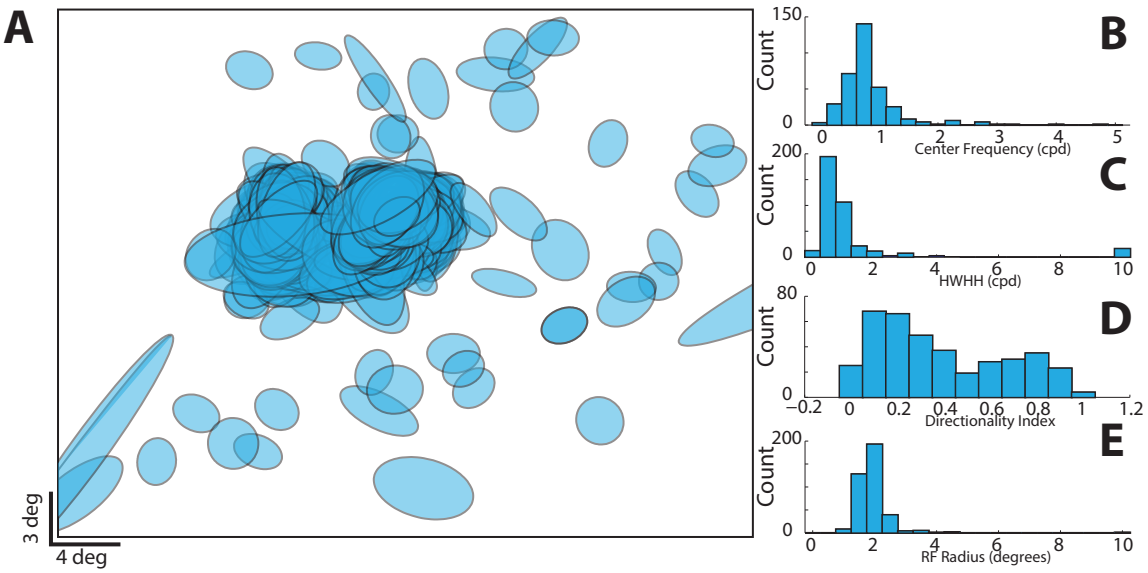

Figure 5.4. Summary of Feature Statistics Across All Channels. A. All receptive field areas indicated in overlapping ellipses (see Methods). Each ellipse captures the total area spanned across both ON and OFF subfields. **B.** The center spatial frequency of a Gaussian fit to the spatial frequency tuning at the highest-responding orientation. **C.** The HWHH (in cycles per degree) of the Gaussian fit to the spatial frequency tuning at the highest-responding orientation. **D.** Directionality index (see Methods) of all channels. Lower values indicates less selectivity. **E.** Radius of the elliptical fit to the classical receptive field 50% contour.

5.3.2 Planar Tilt Stimulus

The first of the feature-based stimuli is called the planar tilt stimulus. Using the sample frames in Figure 5.5E as a reference, the stimulus represents a normal planar drifting grating (middle image, as would be used to map tuning curves) that has been rotated around an axis within the viewing plane. One can imagine projecting the sinusoidal grating onto an LCD monitor and then pushing one side of the monitor away from your face and simultaneously pulling the other side closer as can be seen with the varying cartoon heads above each sample image; this is the same effect as is captured in the tilt stimulus. From all four cardinal orientations, 9 different tilt angles were presented with both positive and negative angles. It is clear that spatial frequency changes across the stimulus, and that the degree to which spatial frequency changes (the spatial derivative) depends on the angle of the tilt and location within the stimulus.

This planar tilt stimulus was presented in an identical manner as the hallway stimulus, capturing the simultaneous activity of 64 neurons at once. To eliminate any adaptation effects, the different tilt angles were shown in a fixed random order. Example results from a single neuron are shown in Figure 5.5. The ON and OFF receptive fields are shown in Figure 5.5A, demonstrating strongly segregated and well-mapped RFs, which was consistent across almost all neurons captured. Figure 5.5B shows the results of mapping the tuning surface of orientation and spatial frequency, which has been interpolated along the spatial frequency axis. This tuning surface indicates mean firing rate and shows moderate orientation tuning, almost no directional selectivity, and strong spatial frequency tuning. Figure 5.5C is a breakout of the spatial frequency tuning curve, fit with a spline curve as used to interpolate it, at a single orientation (180 degrees; leftward drifting). The black dashed line shows the base spatial frequency with no tilt, while the blue line to the immediate left shows the local spatial frequency (at the center of the receptive field, see Methods) at the

full positive tilt (+45 degrees) and the red dashed line to the right shows the local spatial frequency at full negative tilt. The relevant space for these lines is shown magnified in the inset. The amount of spatial frequency spectrum covered by these two extremes depended strongly on the location of the receptive field, but in practice almost all receptive fields were in the same location and thus the spectrum coverage was nearly identical.

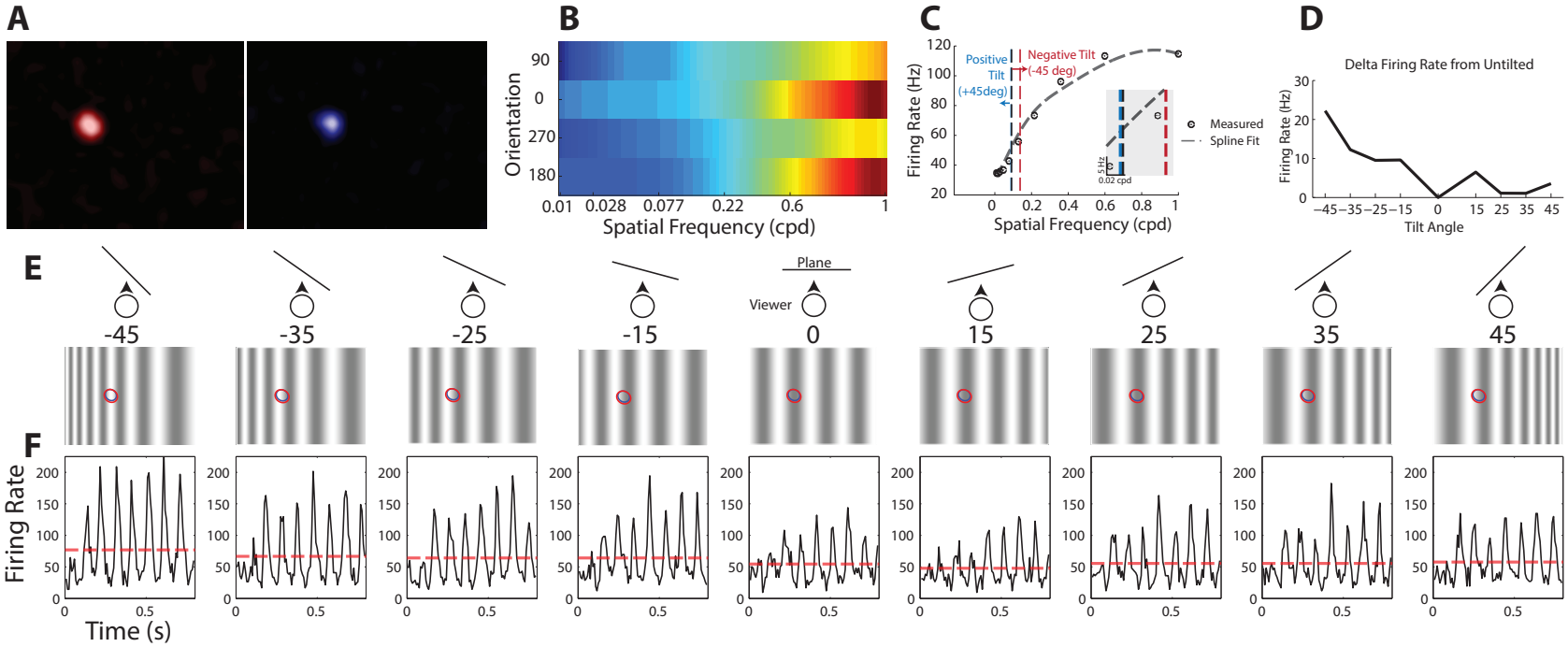

Figure 5.5. Example of Tilt Stimulus Response. A. Receptive field ON (red) and OFF (blue) subfields. **B.** Spatial Frequency-Orientation tuning plot. This neuron is selective for vertical gratings moving left and right, but shows little directional selectivity. **C.** Spatial frequency tuning curve at the orientation for stimuli shown below (leftward) in E. The red bar shows the spatial frequency at the most negatively tilted stimulus while the blue bar shows the spatial frequency at the most positively tilted stimulus. Inset shows a zoom of the relevant area of the tuning curve. In dark gray is a spline fit to the tuning curve. **D.** The mean firing rate (red dashed bars in F) changes as a function of tilt angle. This shows the delta firing rate from the un-tilted stimulus (R_0). **E.** Example frames from each tilt stimulus with receptive field contours (50%) overlaid. **F.** PSTH recorded for each stimulus in **E.** Red dashed bar shows mean firing rate. This neuron was driven at 10 Hz, twice the fundamental frequency of the stimulus.

In Figure 5.5F the sample frames are shown in order with the OFF (blue ellipse) and ON (red ellipse) 50% contours overlaid. This amount of overlap was again consistent across almost all recorded neurons regardless of directional selectivity. Combining our knowledge of the spatial frequency tuning curve results as well as the spatial location of the receptive field, we would generally expect this neuron to increase its firing rate for negative tilts and decrease or have no change for positive tilts. Examining the PSTH of the response across 50 trials in Figure 5.5F, we see that this expectation holds true. Red dashed lines indicate the mean response across all 800 ms of stimulus presentation, and it is clear both the sinusoidal modulations of the response and the mean response rise significantly at -45 degrees tilt and this modulation decreases as the tilt angle decreases. To summarize the results from these individual PSTHs and to compare across neurons which have dissimilar firing rates for un-tilted stimuli (tilt angle of zero), we represent this neuron as a curve of delta firing from zero tilt (reference Figure 5.5D). This analysis method assumes that the expected mean firing rate one would predict from the tuning curve for identical drifting grating parameters as the no tilt stimulus matches the observed response under no tilt conditions. Further, because we seek to quantify only how much the response changes compared to how much we would have predicted it would change, plotted are only absolute values for the delta response (see Methods).

5.3.3 Measuring Tilt Response

How do we represent an expected firing of this neuron? Similar to our encoding strategy for the hallway, we use the knowledge of the local spatial frequency and orientation, combined with our tuning surface map, to generate an expected mean firing rate for a given tilt stimulus. This procedure is shown diagrammatically in Figure 5.6. Different angles of tilt generate different mean local spatial frequencies within the receptive field (marked by each color in Figure 5.6A), which in turn map

to different expected firing rates. If we then translate these mean firing rates onto a curve that organizes them by tilt angle (Figure 5.6B), we can predict what the expected delta firing rate would be for a given neuron (dashed gray line in Figure 5.6C). Changes in tilt angle lead to monotonic changes in local mean spatial frequency and pairing this with a smooth approximation to the spatial frequency tuning curve yields a smooth and monotonic (but nonlinear) relationship between tilt angle and firing rate. Comparing this to the delta response from un-tilted that was actually observed for this neuron (repeated from Figure 5.5D), it is clear that in this case we observed a much higher change than was actually expected (by over two-fold) at -45 degrees of tilt.

5.3.4 Population Measurements of Tilt Response

Repeating this same delta firing rate from un-tilted for all neurons in a given simultaneously recorded group and stimulus orientation, we can attempt to capture this observation more generally. Observed and expected delta from un-tilted responses are shown for two different stimulus orientations in Figure 5.7A, C. Gray lines illustrate individual neurons while solid black lines illustrate the average of all gray lines. Compared between observed (top) and expected (bottom) responses, we can see that there appear to be errors in the expected change from un-tilted. The degree of this significance is illustrated in Figure 5.7B, D; for most tilt angles the observed firing is significantly different ($p < 0.001$, Wilcoxon signed-rank test) from the expected firing rate. This effect is most magnified at the -45 degree tilt angle. This effect is summarized across all simultaneously recorded sets of neurons as well as all orientations in Figure 5.8, noting that it is most magnified at -45 degrees due to the location of the receptive fields. To do this analysis, we captured only the maximum difference between observed and expected delta curves; for both of the examples in Figure 5.7 this would be the value at -45 degrees. This was done to reduce effects

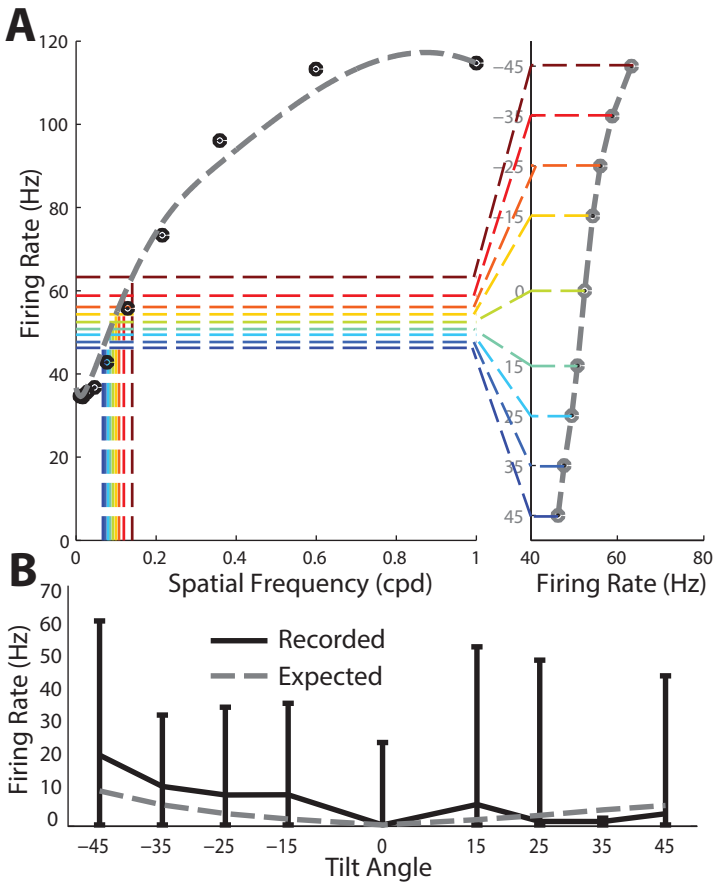

Figure 5.6. Generation of an Expected Tilt Stimulus Response. A. Repeat of the spatial frequency tuning curve from Figure 5.5. Dashed lines indicate the different tilt angles and their associated local spatial frequency (see Methods). These lines are projected to show which tilt they associate with and how this leads to a firing rate prediction. **B.** The change from un-tilted responses as a function of tilt angle are shown for both the recording and the prediction. The prediction is significantly different from the recorded. Lines show the minimum and maximum values observed for each tilt angle.

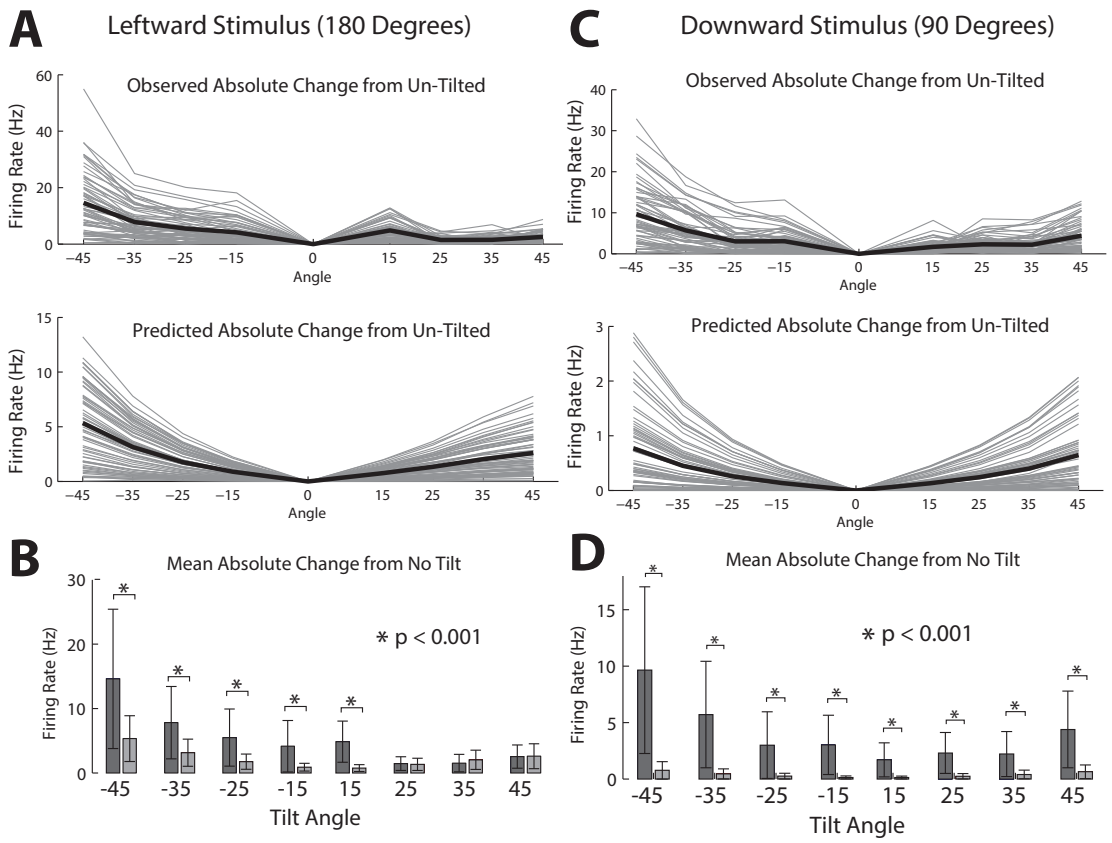

Figure 5.7. Example results from across a population of neurons for two stimuli. A. Observed (top) and Predicted (bottom) change from un-tilted for all neurons in one simultaneously recorded group to the leftward stimulus. **B.** At most tilt angles (0 excepted, since there is no variance) the observed change is significantly higher ($p < 0.001$, Wilcoxon signed-rank test) than was predicted from a tuning-curve driven model. **C.** Same as A, but for a downward stimulus (orthogonal to A). **D.** As in B, but every tilt angle has a significantly higher observed change than was predicted ($p < 0.001$).

caused by differences in orientation tuning and receptive field location. In this we see that across all conditions, the estimated firing rate change undershoots by 3 Hz and this difference remains strongly significant ($p < 0.001$, Wilcoxon signed-rank test). In every case the maximum ΔR_{theta} predicted is 2 to 3 times less than the observed value, with some simultaneous groups having a 10 Hz difference.

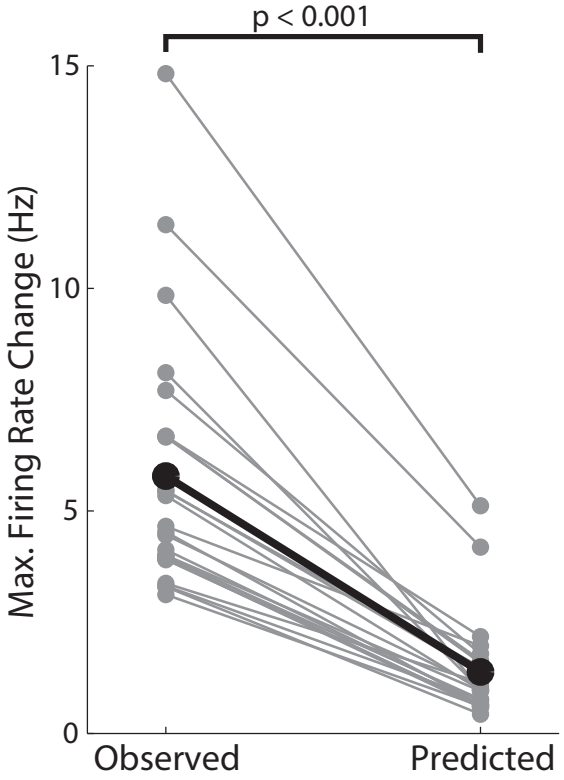

Figure 5.8. Summary of Firing Rate Changes. Each point shows the average maximum expected change in response from un-tilted for a given simultaneously recorded population and stimulus, paired with the predicted results from the same. Black points show average across all gray lines. The overall average observed maximum change is significantly different from the predicted maximum change ($p < 0.001$).

5.3.5 Decoding Tilt Angle

These observed errors of approximately 3 Hz, on average, do not at first glance appear to be an obvious cause for distinction between observed and predicted properties. To more accurately capture how specific tilt angles actually affect the neural response, we built a maximum likelihood decoder that attempts to classify single trial responses into the correct tilt angle stimulus. By analyzing the different tilt angle responses in this manner the way in which non-fixed spatial frequency actually modulates stimulus information becomes more clear, and the results can be seen in Figure 5.9. Single trial responses were chosen from one of each of the 9 different tilt angles presented and single channels as well as the joint population provided estimates of which tilt angle the single trial was chosen from. This process was performed for both the observed responses (i.e. driven by recorded single trials) as well as single trials that were drawn from predicted responses from tuning curve estimates (see Methods for a more in-depth description).

Comparing the performance of our decoder that is built on observed responses to different tilt stimuli and one that is built upon feature selectivity captured by tuning curves (through local spatial frequency changes caused by different tilt stimuli), it is clear that the observed responses have vastly greater differentiation between stimuli than would be expected from the tuning properties. This reflects the significant error in predicting firing rates we have consistently seen in this analysis. Moreover, these results make it clear that there is a significant amount of information about stimulus tilt contained within the neural response to these stimuli and this information is completely discarded when tuning properties fail to reflect this sensitivity.

5.3.6 Motion Boundary Stimulus

We have shown that tilt stimuli can potentially account for at least some of the observed encoding errors but what effect do motion boundaries (boundaries between

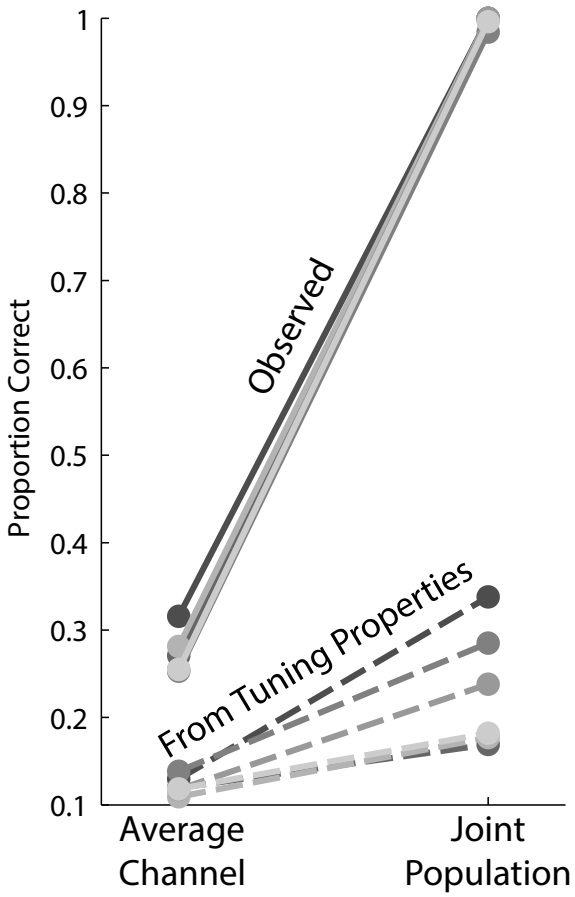

Figure 5.9. Decoding Stimulus Tilt Angle. A decoder of stimulus tilt angle (based upon classifying which stimulus had been presented) provides fairly poor single-trial estimates but almost 100% joint population estimates. A decoder instead based upon local spatial frequency caused by different tilt angles, paired with observed distributions from the tuning mapping process, provided vastly greater inaccuracy.

areas of vastly different orientation or speed) have on neural responses? Motion boundaries have been found to be important in natural scenes, with Roth and Black (2007) finding them highly prevalent in multiple sets of natural scenes. Motion boundary responses have been observed in multiple visual cortical areas (Larsson et al., 2010), yet their importance for primary visual cortex is unclear. The idea of motion boundaries does bear some resemblance to the analysis of non-classical surround suppression effects in V1 (Jones et al., 2001) in that we are looking at changes in neural response caused by changes in orientation that may occur within or outside of the classical receptive field. Furthermore, how important is the location of this motion boundary to the response of an individual neuron? So far all of the analysis presented has been predicated on the stimulus properties present inside the classical receptive field. We can measure the size of this receptive field somewhat arbitrarily by drawing contours that represent different sizes and levels of these receptive fields, yet each of these different sizes changes how we interpret how the stimulus elicits responses in the neuron. By analyzing these shifting visual elements within the stimulus space, we can estimate how nonlinear natural elements affect the responses even outside the classical receptive field.

To address this we created an additional unique stimulus sequence, composed of two orthogonally oriented but otherwise identical sinusoidal drifting grating motion fields, with example frames shown in Figure 5.10. The boundary between these fields was varied between 17 locations, also measuring the response to full-field representations of each drifting grating for a total of 19 stimuli. Two versions of each stimulus were presented, one of which was compressive (where the orientations pointed towards the motion boundary) and one of which was expansive (orientation pointed away from the motion boundary). The response of an example channel to these stimuli is shown in Figure 5.10, where it is clear that the amplitude of the modulations depend somewhat on the location of the motion boundary, but the mean response (red dashed

line) varies rather little.

5.3.7 Location-Independent Measurement of Motion Boundary Distance

To represent the recorded responses in a manner that is independent of receptive field location or size, we first calculated the minimum distance between the center of the recorded receptive field and the motion boundary (see Methods and Figure 5.11). This distance could be negative, in which case the boundary was above the receptive field (rightmost panel), or positive and the boundary was below the receptive field (leftmost panel). This distance was then normalized by the radius of the neurons receptive field. In doing so the different motion boundary stimuli are indexed individually for each neuron by how many receptive field radii away the motion boundary actually was; values less than 1 indicate the motion boundary is inside the classical receptive field.

It was also necessary to generate an expected response based on the location of the motion boundary, similar to what was done for the tilt stimuli. In this case we used a process identical to the receptive field area weighted model used for encoding (see Methods). The predicted response is fixed for 100% of one motion field and 100% of the other at the observed response for these full-field gratings. Once the motion boundary entered the classical receptive field, the expected response would transition smoothly between these two fixed values, depending on the proportion of each motion field within the classical receptive field. This analysis tells us how well any linear model (based either on turning curves or not) would be expected to capture the response to different locations of motion boundaries.

For a number of example neurons we can see the results of this analysis in Figure 5.12. The blue curve shows the measured mean response as a function of motion boundary distance (in arbitrary units of RF radius), with the 95% confidence interval of that mean shown in shaded blue. The red curve shows the expected mean firing rate, with the red shading showing the same 95% confidence interval. In all cases

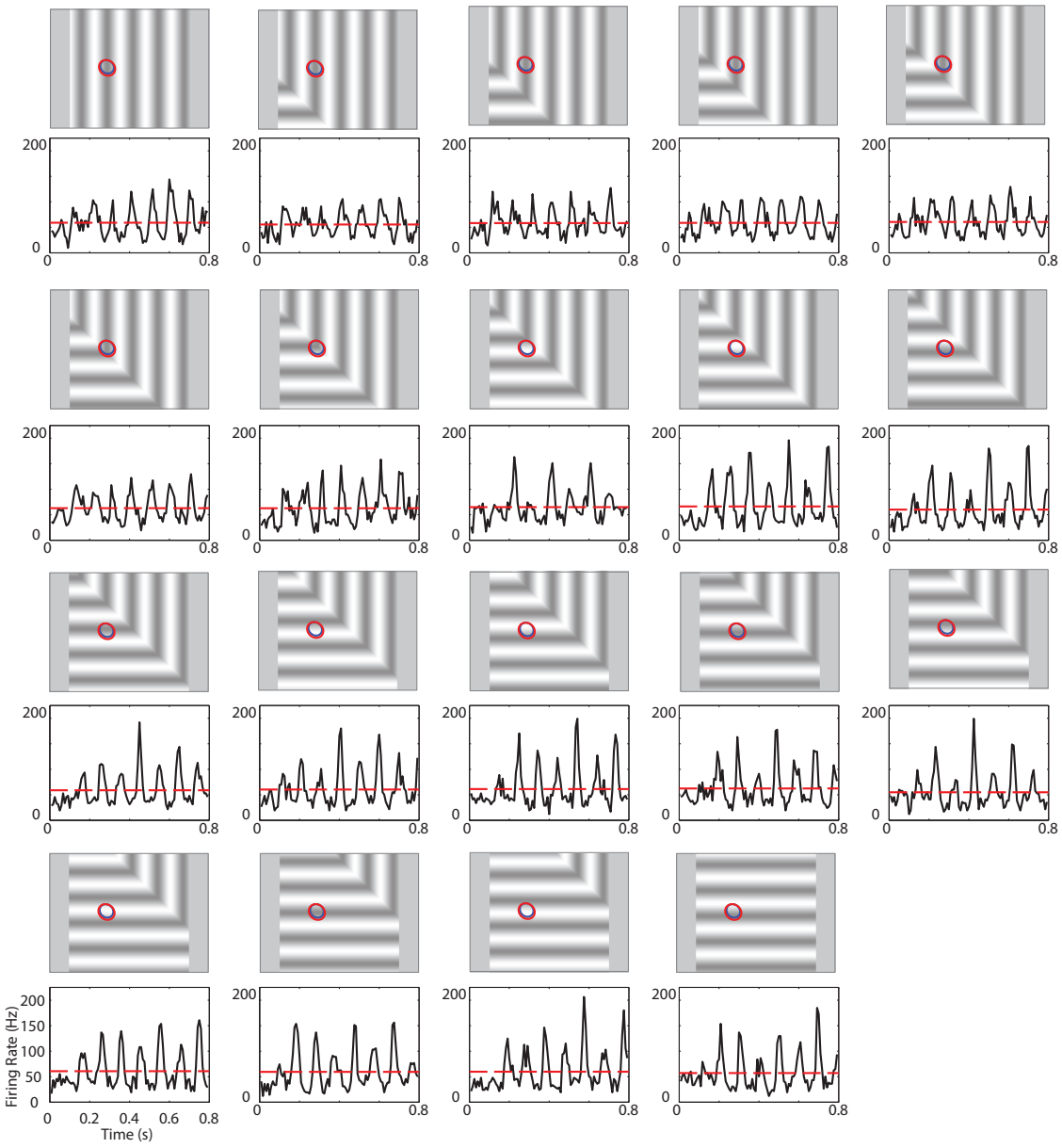

Figure 5.10. Sample Motion Boundary Response. For the same neuron as in Figure 5.5 the response to each of the distinct motion boundary location stimuli is shown. Red dashed bar shows mean response for each PSTH. Although the mean changes very little between cases, it is clear that when the boundary passes through the receptive field there is a switch between two different types of response amplitude.

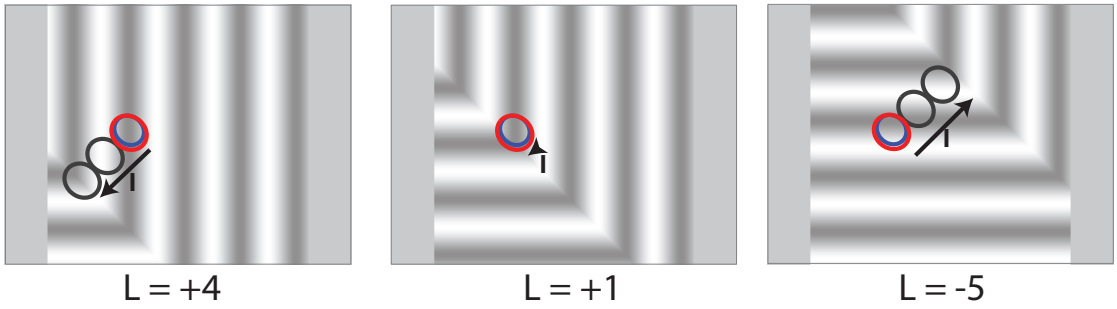

Figure 5.11. A Location and Size Independent Measure of Motion Boundary Location. Distance between the neuron's receptive field and the boundary between unique motion fields was computed as the shorted perpendicular distance scaled by the radius of the receptive field, thus we can estimate how far outside the classical receptive field activation occurs. Here we illustrate this graphically by repeating receptive fields until they reach the motion boundary and count how many radii it takes.

we can see that the prediction actually captures the observed responses reasonably well, with a few errors. The magnitude of these errors passes past the confidence interval only when the motion boundary is near the classical receptive field. Given this observation it would be better to have had more stimuli that were tightly packed within the limited space where the motion boundary passes through the receptive field center.

5.3.8 Population Measurements of Motion Boundary Effects

Taking the error between the blue and red curves and averaging this error across all neurons within a group and for identical stimulus class (compressive or expansive), we find that the error depends strongly on the location of the motion boundary (Figure 5.13B). For all populations in the expansive case there is a peak in error between 0 and 5 receptive field radii and for all populations in the compressive case there is a wider peak between -10 and +5 receptive field radii, indicating that within proximity to a neurons receptive field motion boundaries, there are deviations from classical linear predictions. Since the error is most prevalent only within the space where the motion boundary is close to the center of the receptive field this suggests a potential

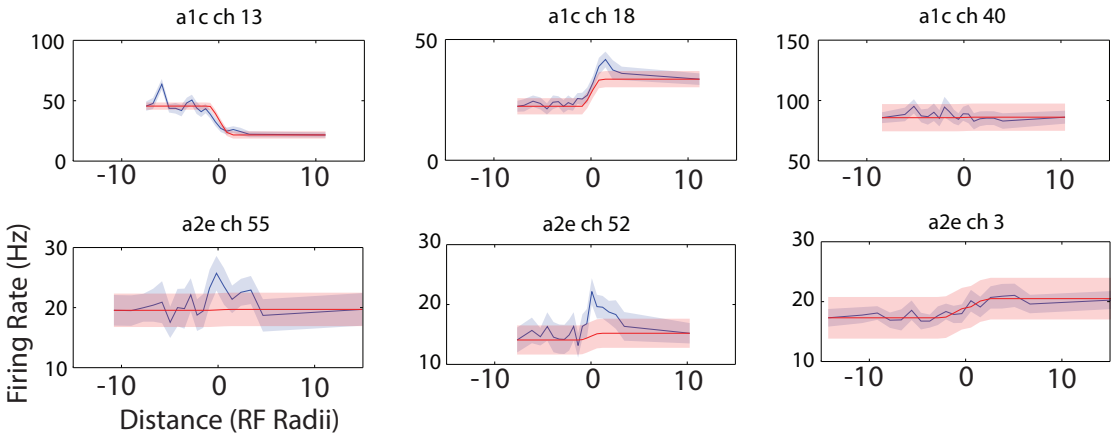

Figure 5.12. Comparisons Between Observed and Expected Motion Boundary Responses. An array of different types of observed relationships between the distance-normalized observed and expected responses. The shaded spaces show 95% confidence intervals of the dark lines that define the estimated mean responses.

model adjustment: where two fields are affecting the space within a few receptive field radii of the center of the receptive field there should be a potentiation of the response relative to the ratio of motion the motion fields. If $\frac{\alpha_{\theta_1}}{\alpha_{\theta_2}} \approx 1$ then this potentiation is maximized

5.3.9 Motion Boundary Classification Decoding

In the same vein as classifying presented tilt stimuli based on the observed responses across single channels and populations, we can also classify different motion boundary locations. Theoretically this is a more challenging classification process because the observed response depends more strongly on the location of the receptive field, but as before we do not adjust for this given how close each recorded receptive field is to all other receptive fields. This proximity means that, for the most part, the motion boundary is equidistant from all receptive field centers at once (with some variation). Moreover it is clear from responses in Figure 5.12 that we would not expect high classification performance for all possible motion boundary locations; even in the view of non-classical receptive field effects, many of the locations should

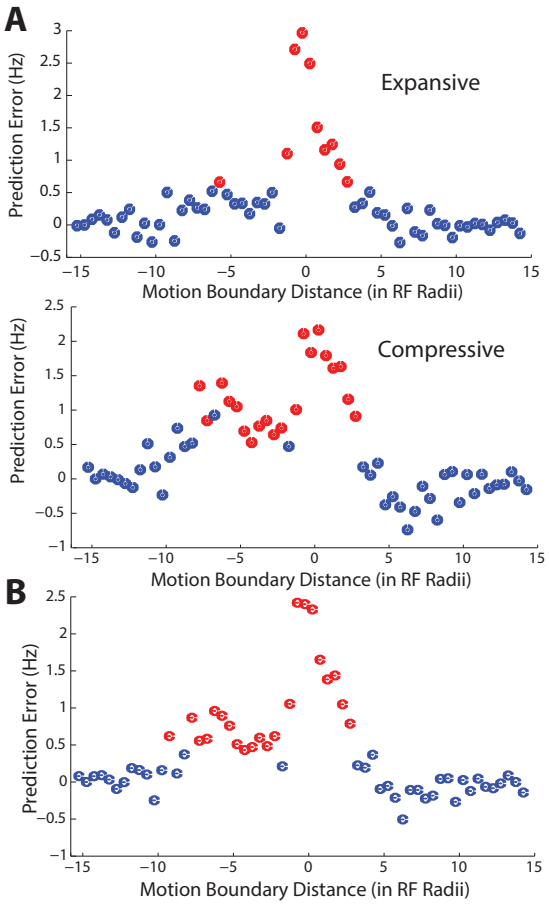

Figure 5.13. Motion Boundary Results. A. Across all neurons within a simultaneous recordings the average prediction error (difference between red and blue curves in A) reaches a small but substantial peak between -5 and 5 receptive field radii away from center of the receptive field. The location of the peak is consistent across expansive and compressive presentations, but what happens at between -5 and 0 receptive field radii varies between the two. The color of the dots indicates significance of the difference between the observed and expected responses for each location (red $p < 0.001$, Wilcoxon signed-rank test). **B.** Across all stimuli the same trend remains, with significant differences between -10 and +5 receptive field radii.

simply appear identical to full-frame grating stimuli. To simplify the classification task in light of the importance of the location of the boundary to the center of the receptive field, the task has been restricted to a discrimination from among 5 possible stimuli rather than all 19 possible stimuli. The five chosen represent each full-frame drifting grating, versions that have 25% of each motion field (boundary in bottom left or top right), and the version that has 50% of the frame taken up by each motion field. This represents a good cross-section of possible boundary locations, and also respects the statistically significant areas within the findings of Figure 5.13.

What we find for these classification results mirrors what we found for the tilt stimuli in Figure 5.9. The average channel performs relatively poorly at the task, again besting chance performance by only two-fold. As before, an estimate based upon the joint probabilities across an entire simultaneously recorded population performs significantly better, with correct classification 100% of the time for all such populations. As before, a classifying decoder was also built using theoretical results from the tuning curve. Each motion boundary stimulus for each neuron has some portion of the receptive field that is in one motion field or another, due to different orientations in each field this may alter the neural response. By creating a distribution of observed responses based upon the underlying proportion of each motion field within the classical receptive field (see Methods), we can approximate a response distribution to the motion boundary stimulus directly from the tuning properties. Just like a similar tuning-driven approach for the planar tilt stimulus, we see that such a population, even at best, has significantly lower classification performance than the true recorded channels.

These decoding results across both the tilt and motion boundary stimuli indicate that there is significant embedded information about perspective and differing object motion that is built into the responses across populations of neurons in primary visual cortex. Classical models of these neurons, relying on tuning properties, fail to

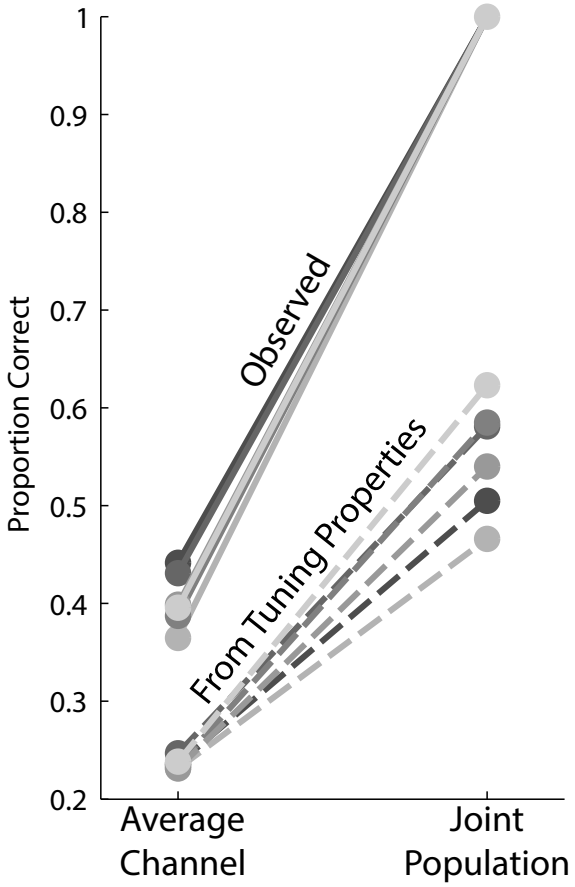

Figure 5.14. Motion Boundary Classification. Similar to the results in Figure 5.9 motion boundary stimuli segments can also be classified. In this case since there are 19 total stimuli we have restricted the set to 5 possible choices: each full-frame stimulus as well as 25% of the horizontal grating with 75% of the vertical grating, half of the field taken up by each grating, and 75% horizontal with 25% vertical. The observed performance is very high still, reaching 100% correct classification for all neural populations when a joint estimate is produced. The performance of a theoretical classifier based upon tuning properties performs better than for the tilt stimuli (due in part to the decreased number of choices) but still performs significantly worse than the actual decoding performance.

capture these population dynamics if the set of tuning properties is not comprehensive enough. More accurate models require some estimate of the tuning to more natural-based effects to fully capture population representations of motion.

5.4 Discussion

5.4.1 Summary

In Chapter 3, we explored the way motion information is represented in populations of neurons in the primary visual cortex. Ultimately we saw that there were challenges to prediction of motion, but it was not clear from where these challenges originated. Based on the actual stimulus phenomena present in the visual hallway stimulus and more generally in natural scenes as a whole, we restricted our investigations to two unique natural-like visual features, each presented along a continuum and independent of each other. The first of these stimuli, called the planar tilt stimulus, simulates the tilt of a sinusoidal grating within the viewing window such that half of the plane moves towards the viewer and the other half moves away such as perspective might create. We found that when the tilt increases local spatial frequency and local spatial frequency gradient, predictions from linear models fail to capture the observed responses and have a significant error. The second stimulus placed two orthgonally orientation fields of drifting gratings within the visual image and moved the location of the boundary between these two fields between many different locations. We found that when the location of the motion boundary is close to (and within) the classical receptive field, traditional linear models again failed to capture the observed response. Across both of these two stimuli, classification of different stimuli (which tilt angle or motion boundary location from a set of possible choices) from the observed responses was significantly more accurate than a hypothetical classifier based on the set of estimated responses from linear models. This suggests that motion information and natural feature information is embedded in the responses of populations of neurons in

a way that classical models of the style presented here with limited tuning selectivity fail to represent.

5.4.2 Similarity to Previous Stimuli

Using either the tilt gratings or the motion boundary gratings, we have created distillations of specific phenomenology in natural scenes. Being such distillations, they have some similarities to stimuli used in previous visual experiments (Liu et al., 2004, Rosenberg et al., 2013, Larsson et al., 2010, Sanada and Ohzawa, 2006), yet in each of these previous studies the focus has been either on binocularity or higher-level locations in the visual pathway. The closest in spirit is Sanada and Ohzawa (2006) in which they explicitly explored a three-dimensional slant of a drifting grating, but in this case the focus was on binocular rivalry and a model that implements detection of spatial frequency differences across each eye. The effects of natural-like features on single eye, single neuron responses in V1 and on population codes therein is so far unexplored.

5.4.3 Traditional Binocular Disparity

The tilt stimulus bears similarity to the stimuli used to investigate binocular disparity and is mostly clearly related to Sanada and Ohzawa (2006). In this study the ability of binocular neurons (and populations thereof) were used to test discriminability of different grating tilts, much like our own stimulus. Sanada and Ohzawa conclude that early visual cortex (i.e. V1) is a potential place for encoding of three-dimensional surface orientations to emerge. This is consistent with our own findings, but we further assert that this information is present monocularly. There did not appear to be a correlation between eye preference and the effect of tilt angle, although our stimulus sets were not made to allow for measuring such preferences easily. Without any binocular information, however, joint populations were easily able to decode different stimulus tilt angles which implies that such information is unnecessary for this kind

of discrimination, at least with artificial types of stimuli. Further investigation in the links between binocular disparity and monocular information about stimulus tilt (and perspective) are likely to provide unique perspectives on how motion and depth information exist within monocular and binocular streams of visual information.

5.4.4 An Axis of "Natural"

These results point to the beginnings of a possible functional definition and also a way in which we can define some continuous axis between artificial and natural scenes: just as the degree of tilt or the proximity of motion boundary modulates how well a linear model would predict the response, it is possible that these (and other) features can be distilled into an effective discriminator of how natural a scene might be classified. Looking just at the responses in Figure 5.5 and Figure 5.7, it is clear that in many cases it is only when spatial frequency has a high local first derivative (i.e. the spatial frequency at one end of a receptive field is noticeably different from the spatial frequency at the other end) that our classical linear tuning-curve driven models fail to capture the neural response. We might hypothesize then that one metric by which to judge natural scenes is the local first derivative of spatial frequency; low or zero values would be artificial scenes and high values would be natural scenes with the continuum between them defining an axis of "natural-ness" of the visual scene in question. Further work would employ both existing and novel datasets for which we can calculate this first derivative of spatial frequency and observe how well existing models predict responses to the scene, as well as qualitatively classify how to order these stimuli based upon how natural they "feel".

5.4.5 Location Variability

We designed our stimulus and the spatial location of the receptive fields to capture highly dense information about the effect of both the tilt angles and motion boundary distances. Due to this design, almost every single one of our receptive fields for the

recorded neurons was in nearly the same location, with small location variations inside a roughly 1 degree radius area. While a wider spread of receptive field locations might have provided more information about how location and local properties generated the observed effects, the complexity of even these tilted or two-field stimuli required the dense sampling. Although these results make it clear that there is an effect due to these natural feature stimuli that is not predicted by the tuning sensitivities we have measured here, it is possible that we would not be able to capture the variations of neuron properties necessary to derive a model that fully explains the effects here.

5.4.6 Features and Stimulus Information

Classical observations of visual neurons, based only on tuning properties, would ascribe no predictive power to natural perspective-like affects or motion boundaries to modulations in the neural response to drifting grating stimuli. Measuring only a few feature selectivities such as orientation or spatial frequency tuning implicitly ignores the underlying relationship between multiple parameters. Like many system identification problems, the accuracy of the system model depends on capturing a high enough dimensionality to explain the majority of the signal variance and our results show that orientation tuning, spatial frequency tuning, and even probably temporal frequency tuning fail to capture enough dimensions. Our results suggest that, at a minimum, when it comes to natural stimuli, such tuning approximations are insufficient to explain neural responses without accounting for nonlinear features as presented here. Although we did not measure such tuning in a way that allows us to propose modifications to the classical models that account for such response modulations, the fact that our classification tasks in both cases perform significantly better than control tuning-driven classifications implies strongly that information about planar tilt and motion boundaries (at a minimum) is in fact embedded within the population response. Tapping in to this information in future experiments could

provide dramatic insight into the way responses to natural scenes differ significantly from what our existing models and understanding of the visual system would expect.

CHAPTER VI

DISCUSSION

6.1 Summary

In this thesis I have explored how different aspects of the visual system transmit information about motion in the outside world. Starting in the LGN I examined how populations of neurons with strong timing correlations transmit information about motion that results in orientation selectivity in their common cortical targets. This provides the first organized general representation of motion within neural responses. Interestingly the resultant orientation tuning captures typical tuning properties and expresses an information transmission efficiency at levels of timing correlation that correspond to those measured from real pairs of neurons. Following this I moved up to cortex and explored a novel natural-like stimulus to explore motion representation in V1. The stimulus was a sinusoidal hallway in which drifting gratings were mapped to the walls of a hallway in 3D projection with active exploration. Across a dense array of recorded V1 neurons I found that stimulus properties such as orientation and spatial frequency are poorly encoded by classical models of tuning under these natural-like conditions. Despite this, decoding of motion classes is accurate when the joint population activity generated estimates of stimulus motion. Finally I explored separated versions of the natural-like features present in the hallway stimulus and found that both spatially varying spatial frequency and motion boundaries both contribute to observed errors in classical encoding models. These observed differences resulted in significantly better decoding of stimulus sequences of both types of stimuli, indicating that both of these feature types induce unique information about stimulus motion that classical models do not capture. All together these results show that

motion information within stimuli both artificial and natural is required to be transmitted by populations of neurons for effective transmission and that populations are required to understand the intricate responses to small natural stimulus features.

6.2 Philosophy of Information Transmission With Single Neurons and Populations

Despite demonstrating strong feature selectivity can result from arrangements of populations of neurons with synapses on a single common target, there is also some apparent doubt as to the useful functionality of such feature selectivity in everyday scenes. The thalamic populations in Chapter 2 generated orientation selectivity under artificial stimulus conditions using drifting sinusoidal gratings and it is not clear from the results presented here what their synchronization would produce under conditions of more natural stimulation. Regardless of how well our synchronization models predict feature selective responses or how well our common linear models capture the information embedded in neural responses it is clear that the information must be encoded in the responses. Even if it is the case that it is poorly encoded in single neurons, as very well may be, it is also apparent that in general significantly denser sampling is necessary to perform decoding tasks that attempt to estimate exact parameters. Just as I hypothesized that feature selectivity found under artificial conditions may bear diminished relevance under natural viewing conditions, it is also possible that feature selectivity as measured was simply a deprived description of a visual neuron. If phenomena like orientation tuning are only part of some more complex tuning or representation of motion within the visual cortex, that would explain why we have error-prone parameter decoding but nearly perfect classification. In this view decoding and classification success depends strongly on what we think the "information" is within the system and the appropriateness of our existing models depends entirely on how accurate this concept of information is.

6.3 Anesthetization and Synchronization

All experiments shown in this work were performed under anesthetization. This is particularly important to consider for our findings here given that previous research has identified anesthesia as a potential modifier of thalamocortical signaling (Stoelzel et al., 2009, Castro-Alamancos, 2002) with unknown effects on synchronization within such circuits (Wang et al., 2010b). Along the same lines, within the primary visual cortex studies in awake rats has revealed markedly different activity from anesthetized including changes in population synchronization (Greenberg et al., 2008). It is clear then that some or all of the results reported here could potentially be the effect of anesthetization (or even worse small changes in the level of anesthesia within an experiment), however even in the awake cat differences between responses to drifting gratings and natural scenes has been reported (Kayser et al., 2003). Thus it is not clear how our actual results would be affected by anesthesia and it is important to further study this.

One thing to consider in light of questions regarding anesthesia is the effect of bursting on the observed responses here, particularly as it relates to arguments about synchronization. Bursting, loosely defined, is the output of several spikes from a single neuron with very short ISIs. It is known that bursting is of value to encoding responses to both artificial and natural scenes (Reinagel et al., 1999, Lesica et al., 2006) and for our suggested mechanism of feature selectivity origination bursting has a special importance. The proposed model (based on historical models as far back as Hubel & Wiesel in 1962) implies that the temporal correlation between spike times in neurons that compose a particular sub-population is of primary importance to determining output feature selectivity. This correlation results from the nonlinear thresholding mechanisms driving membrane voltage but is driven by an assumed linear contribution from within the population which drives the input to that membrane voltage. Under bursting schemes, however, particular neurons within that population

might begin to overwhelm the action of the population as a whole; all other things being equal an input neuron in a burst mode will more strongly drive integration than an input neuron with consistent firing, as synchronization will be more strongly linked to the burst spikes than the tonic ones. If our anesthetized data fails to represent true bursting rates, which we cannot estimate due to the choice to use multiunit data for visual cortex and partially simulated data for LGN, then we could also fail to capture the true importance of bursting in determining our synchrony-driven feature selectivity. Indeed previous research indicates that bursting is significantly more common in anesthesia than it is in awake animals (Alitto et al., 2011). This presents at least the possibility that some of our observations as they concern synchronization could be affected by this over-prevalence and events we attribute solely to synchronization might be driven by synchronization AND bursting. This view is tempered though by a qualitative analysis of the responses in Figure 3.3, although neurons do show some protracted silence at the 90 degree condition the spiketimes do not cluster into particularly apparent chunks of burstiness. Further research will clarify the importance of this observation.

6.4 Future Directions

Further experiments are necessary to continue to expand the scope and clarity of the arguments presented here. Most pressing is an experiment designed to bridge the results between the chapters in a conclusive manner and to demonstrate positively that bottom-up feature selectivity remains valid and appropriate when presented with natural scenes. At stake primarily is the kind of disconnect that can be potentially interpreted in the focus of the works in Chapters 2 and 3; first the focus is on collaborative and synchronous information transmission from a population to a single neuron and the second focuses on errors in single-neuron response prediction. Concluding, we show that the population of neurons within the primary visual cortex,

however, contain sufficient information within at least some measure (either decoding or classification) to have different responses to stimuli with different motion effects. Unexplored here is how well these populations can then transmit this information forward to a hypothetical next stage of processing with single neurons (i.e. Area MT, to which V1 projects). The primary visual cortex is hardly the "end of the line" for visual transmission and so it remains logical that despite having significantly more elegant feature selectivity than the lateral geniculate nucleus, the manner in which it passes specific information forward synchronously should remain nearly identical. This needs to be tested though.

The most elegant experiment to propose, which remains out of reach technologically, would be a hypothetical experiment that permitted simultaneous recording from LGN, V1, and MT for identical receptive fields within visual space. This forms an exquisite 3-layer axis of spatial sampling and synchronization. The ideal experiment would demonstrate that a very large population of LGN neurons synchronously drive a moderate population of V1 neurons, which in turn synchronously drive a handful of MT (or other higher-level) neurons. The benefits of providing such a multi-tiered synchrony-driven model are clear, particularly in the context of the work presented here. First, it is not necessary to have extensively accurate models of individual neural responses, but rather account for the accumulated trial-to-trial spiketimes across entire populations, which restricts the precision necessary to predict individual spike times and PSTH functions. Second it decouples the information transmission and estimation from intermediate steps and eliminates any dependence on artificially measured tuning functions.

With such an experiment and model representation of neural signaling, it will become clearer to what extent natural scenes engage the entire pathway in manners that are both consistent and different from classical expectations. It will also free our understanding of the pathway from stopgap abstractions the represent the individual

stages; instead of having to represent each layer as being composed of individual neurons with tuning functions of potentially large dimensionality, everything we need to know about the stimulus and it's motion (in this case) is derived from relatively simple low-level visual receptive field representations and the intricate synchronization relationships between populations. Such representations could ultimately form model systems for all sensory processing.

6.5 Final Conclusions

Natural scenes in the context of Vision Neuroscience provide a phenomenally exciting realm of expansion to the classical abstract visual stimuli. While it is tempting to assume that lessons about the way visual neurons represent the outside world transition easily from one stimulus type to the other and thus we can break down single features and learn how they drive visual neurons, it does not appear to be the case that things are so easy. Rather than building on top of our understanding of classical tuning properties, it appears as though some inherent aspect of natural scenes fundamentally alters the way these visual neurons and populations actually represent these tuning properties. With this in mind it is important to study visual phenomena under both artificial and natural conditions to understand how well the principles generalize to daily (and useful) visual scenarios.

Bibliography

Abbott, L. F. and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101.

Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America. A, Optics and image science*, 2(2):284–99.

Alitto, H. J., Moore, B. D., Rathbun, D. L., and Usrey, W. M. (2011). A comparison of visual responses in the lateral geniculate nucleus of alert and anaesthetized macaque monkeys. *The Journal of physiology*, 589(Pt 1):87–99.

Alonso, J. M., Usrey, W. M., and Reid, R. C. (1996). Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, 383(6603):815–9.

Alonso, J. M., Usrey, W. M., and Reid, R. C. (2001). Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *Journal of Neuroscience*, 21(11):4002.

Anderson, J. S. (2000). The Contribution of Noise to Contrast Invariance of Orientation Tuning in Cat Visual Cortex. *Science*, 290(5498):1968–1972.

Anderson, J. S., Lampl, I., Gillespie, D. C., and Ferster, D. (2001). Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(6):2104–12.

Averbeck, B. B. and Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *Journal of neurophysiology*, 95(6):3633–44.

Azouz, R. and Gray, C. M. (2003). Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. *Neuron*, 37(3):513–23.

Baccus, S. a., Olveczky, B. P., Manu, M., and Meister, M. (2008). A retinal circuit that computes object motion. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(27):6807–17.

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. a., and Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings. Biological sciences / The Royal Society*, 264(1389):1775–83.

Ban, H., Preston, T. J., Meeson, A., and Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature neuroscience*, 15(4):636–643.

Barlow, H. B. (1963). Slippage of contact lenses and other artefacts in relation to fading and regeneration of supposedly stable retinal images. *Quarterly Journal of Experimental Psychology*, 15(1):36–51.

Benucci, A., Ringach, D. L., and Carandini, M. (2009). Coding of stimulus sequences by population responses in visual cortex. *Nature neuroscience*, 12(10):1317–24.

Born, R. T. and Bradley, D. C. (2005). Structure and function of visual area MT. *Annual review of neuroscience*, 28:157–89.

Bradley, a., Skottun, B. C., Ohzawa, I., Sclar, G., and Freeman, R. D. (1987). Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. *Journal of neurophysiology*, 57(3):755–72.

Bridge, H. and Cumming, B. G. (2001). Responses of macaque V1 neurons to binocular orientation differences. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(18):7293–302.

Briggs, F., Mangun, G. R., and Usrey, W. M. (2013). Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature*, 499(7459):476–480.

Bruno, R. M. and Sakmann, B. (2006). Cortex is driven by weak but synchronously active thalamocortical synapses. *Science (New York, N.Y.)*, 312(5780):1622–7.

Buracas, G. T., Zador, a. M., DeWeese, M. R., and Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20(5):959–69.

Butts, D. a., Desbordes, G., Weng, C., Jin, J., Alonso, J.-M., and Stanley, G. B. (2010). The episodic nature of spike trains in the early visual pathway. *Journal of neurophysiology*, 104(6):3371–87.

Butts, D. A. and Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS biology*, 4(4):e92.

Butts, D. a., Weng, C., Jin, J., Yeh, C.-I., Lesica, N. a., Alonso, J.-M., and Stanley, G. B. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature*, 449(7158):92–5.

Campbell, F., Cooper, G., and Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *The Journal of Physiology*, 203(1):223.

Carandini, M. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS biology*, 2(9):E264.

Carandini, M. and Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(1):470–84.

Cardin, J. A., Kumbhani, R. D., Contreras, D., and Palmer, L. A. (2010). Cellular mechanisms of temporal sensitivity in visual cortex neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(10):3652–62.

Castro-Alamancos, M. a. (2002). Role of thalamocortical sensory suppression during arousal: focusing sensory inputs in neocortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(22):9651–5.

Chen, Y., Geisler, W. S., and Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature neuroscience*, 9(11):1412–20.

Chichilnisky, E. and Kalmar, R. (2003). Temporal resolution of ensemble visual motion signals in primate retina. *Journal of Neuroscience*, 23(17):6681.

Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature neuroscience*, 14(7):811–9.

Coppola, D. and Purves, D. (1996). The extraordinarily rapid disappearance of entopic images. *Proceedings of the National Academy of Sciences of the United States of America*, 93(15):8001–4.

Cumming, B. G. and Parker, A. J. (1999). Binocular Neurons in V1 of Awake Monkeys Are Selective for Absolute, Not Relative, Disparity. *J. Neurosci.*, 19(13):5602–5618.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982a). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5):545–59.

De Valois, R. L., Yund, E. W., and Hepler, N. (1982b). The orientation and direction selectivity of cells in macaque visual cortex. *Vision research*, 22(5):531–44.

Deneve, S., Latham, P. E., and Pouget, a. (1999). Reading population codes: a neural implementation of ideal observers. *Nature neuroscience*, 2(8):740–5.

Derrington, A. M. and Fuchs, A. F. (1979). Spatial and temporal properties of X and Y cells in the cat lateral geniculate nucleus. *The Journal of physiology*, 293:347–64.

Desbordes, G., Jin, J., Alonso, J.-M., and Stanley, G. B. (2010). Modulation of temporal precision in thalamic population responses to natural visual stimuli. *Frontiers in systems neuroscience*, 4(November):151.

Desbordes, G., Jin, J., Weng, C., Lesica, N. A., Stanley, G. B., and Alonso, J.-M. (2008). Timing precision in population coding of natural scenes in the early visual system. *PLoS biology*, 6(12):e324.

Duffy, C. J. and Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of neurophysiology*, 65(6):1329–45.

Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 23:441–71.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America. A, Optics and image science*, 4(12):2379–94.

Finn, I. M., Priebe, N. J., and Ferster, D. (2007). The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex. *Neuron*, 54(1):137–52.

Fleet, D. J., Wagner, H., and Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, 36(12):1839–1857.

Frechette, E. S., Sher, A., Grivich, M. I., Petrusca, D., Litke, A. M., and Chichilnisky, E. J. (2005). Fidelity of the ensemble code for visual motion in primate retina. *Journal of neurophysiology*, 94(1):119–35.

Geisler, W. S. and Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Visual neuroscience*, 14(05):897–919.

Gerstner, W. and Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, Cambridge, UK.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin Company, Boston.

Gizzi, M. S., Katz, E., Schumer, R. A., and Movshon, J. A. (1990). Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *Journal of neurophysiology*, 63(6):1529–43.

Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–7.

Greenberg, D. S., Houweling, A. R., and Kerr, J. N. D. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature neuroscience*, 11(7):749–51.

Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in psychology*, 4(October):777.

Grunewald, A. and Skoumbourdis, E. K. (2004). The integration of multiple stimulus features by V1 neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(41):9185–94.

Gur, M., Beylin, A., and Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 17(8):2914–20.

Hicks, T. P., Lee, B. B., and Vidyasagar, T. R. (1983). The responses of cells in macaque lateral geniculate nucleus to sinusoidal gratings. *The Journal of physiology*, 337(337):183–200.

Hubel, D. H. (1988). *Eye, Brain, and Vision*. Scientific American Library, New York.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–54.

Ince, R. a. a., Panzeri, S., and Kayser, C. (2013). Neural Codes Formed by Small and Temporally Precise Populations in Auditory Cortex. *Journal of Neuroscience*, 33(46):18277–18287.

Jin, J., Wang, Y., Swadlow, H. a., and Alonso, J. M. (2011). Population receptive fields of ON and OFF thalamic inputs to an orientation column in visual cortex. *Nature neuroscience*, 14(2):232–8.

Johnson, D. H. and Ray, W. (2003). Optimal stimulus coding by neural populations using rate codes. *Journal of computational neuroscience*, 16(2):129–38.

Jones, H. E., Grieve, K. L., Wang, W., and Sillito, a. M. (2001). Surround suppression in primate V1. *Journal of neurophysiology*, 86(4):2011–28.

Jones, J. P. and Palmer, L. a. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1187–211.

Josić, K., Shea-Brown, E., Doiron, B., and de la Rocha, J. (2009). Stimulus-dependent correlations and population codes. *Neural computation*, 21(10):2774–804.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors (2000). *Principles of Neural Science*. McGraw-Hill, New York, 4 edition.

Kara, P., Reinagel, P., and Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3):635–46.

Kayser, A. S. and Miller, K. D. (2002). Opponent inhibition: a developmental model of layer 4 of the neocortical circuit. *Neuron*, 33(1):131–42.

Kayser, C., Salazar, R. F., and Konig, P. (2003). Responses to natural scenes in cat V1. *Journal of neurophysiology*, 90(3):1910–20.

Kelly, S. T., Kremkow, J., Jin, J., Wang, Y., Wang, Q., Alonso, J.-M., and Stanley, G. B. (2014). The role of thalamic population synchrony in the emergence of cortical feature selectivity. *PLoS computational biology*, 10(1):e1003418.

Klam, F., Zemel, R. S., and Pouget, a. (2008). Population coding with motion energy filters: the impact of correlations. *Neural computation*, 20(1):146–75.

KUFFLER, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology*, 16(1):37–68.

Kumbhani, R. D., Nolt, M. J., and Palmer, L. a. (2007). Precision, reliability, and information-theoretic analysis of visual thalamocortical neurons. *Journal of neurophysiology*, 98(5):2647–63.

Larsson, J., Heeger, D. J., and Landy, M. S. (2010). Orientation selectivity of motion-boundary responses in human visual cortex. *Journal of neurophysiology*, 104(6):2940–50.

Lesica, N. a., Weng, C., Jin, J., Yeh, C.-I., Alonso, J.-M., and Stanley, G. B. (2006). Dynamic encoding of natural luminance sequences by LGN bursts. *PLoS biology*, 4(7):e209.

Li, F. F. and Perona, P. (2005). A Bayesian Hierarchical Model for Learning Natural Scene Categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 524–531. IEEE.

Liu, Y., Vogels, R., and Orban, G. a. (2004). Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(15):3795–800.

McLaughlin, D., Shapley, R., Shelley, M., and Wielaard, D. J. (2000). A neuronal network model of macaque primary visual cortex (V1): orientation selectivity and dynamics in the input layer 4Calpha. *Proceedings of the National Academy of Sciences of the United States of America*, 97(14):8087–92.

Millard, D. C., Wang, Q., Gollnick, C. a., and Stanley, G. B. (2013). System identification of the nonlinear dynamics in the thalamocortical circuit in response to patterned thalamic microstimulation in vivo. *Journal of neural engineering*, 10(6):066011.

Miller, K. and Troyer, T. (2002). Neural Noise Can Explain Expansive , Power-Law Nonlinearities in Neural Response Functions. *Journal of neurophysiology*, 87(2):653.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., and Newsome, W. (1983). The Analysis of Moving Visual Patterns. In Chagas, C., Gattass, R., and Gross, C., editors, *Pattern Recognition Mechanisms*. Springer-Verlag, New York.

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *The Journal of physiology*, 283:101–20.

Nadler, J. W., Barbash, D., Kim, H. R., Shimpi, S., Angelaki, D. E., and Deangelis, G. C. (2013). Joint Representation of Depth from Motion Parallax and Binocular Disparity Cues in Macaque Area MT. 33(35):14061–14074.

Neri, P. (2014). Semantic Control of Feature Extraction from Natural Scenes. *Journal of Neuroscience*, 34(6):2374–2388.

Nishimoto, S. and Gallant, J. L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(41):14551–64.

Palmer, S. E. and Miller, K. D. (2007). Effects of inhibitory gain and conductance fluctuations in a simple model for contrast-invariant orientation tuning in cat V1. *Journal of neurophysiology*, 98(1):63–78.

Paradiso, M. A. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological cybernetics*, 58(1):35–49.

Perrone, J. a. (2004). A visual motion sensor based on the properties of V1 and MT neurons. *Vision research*, 44(15):1733–55.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–9.

Pouget, A., Deneve, S., Ducom, J.-C., and Latham, P. E. (1999). Narrow Versus Wide Tuning Curves: What's Best for a Population Code? *Neural Computation*, 11(1):85–90.

Pouget, a., Zhang, K., Deneve, S., and Latham, P. E. (1998). Statistically Efficient Estimation Using Population Coding. *Neural computation*, 10(2):373–401.

Priebe, N. J. and Ferster, D. (2005). Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron*, 45(1):133–45.

Priebe, N. J., Lisberger, S. G., and Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(11):2941–50.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M., editors (2004). *Neuroscience*. Sinauer Associates Inc., Sunderland, MA.

Quian Quiroga, R. and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews. Neuroscience*, 10(3):173–85.

Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., and Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *Journal of neurophysiology*, 77(5):2836–41.

Reichardt, W. (1987). Evaluation of optical motion information by movement detectors. *Journal of comparative physiology. A, Sensory, neural, and behavioral physiology*, 161(4):533–47.

Reid, R. C. and Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378(6554):281–4.

Reinagel, P., Godwin, D., Sherman, S. M., and Koch, C. (1999). Encoding of visual information by LGN bursts. *Journal of neurophysiology*, 81(5):2558–69.

Reinagel, P. and Zador, a. M. (1999). Natural scene statistics at the centre of gaze. *Network (Bristol, England)*, 10(4):341–50.

Ringach, D. L. (2004). Haphazard wiring of simple receptive fields and orientation columns in visual cortex. *Journal of neurophysiology*, 92(1):468–76.

Rose, D. and Blakemore, C. (1974). An analysis of orientation selectivity in the cat's visual cortex. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 20(1):1–17.

Rosenberg, a., Cowan, N. J., and Angelaki, D. E. (2013). The Visual Representation of 3D Object Orientation in Parietal Cortex. *Journal of Neuroscience*, 33(49):19352–19361.

Roth, S. and Black, M. J. (2007). On the Spatial Statistics of Optical Flow. *International Journal of Computer Vision*, 74(1):33–50.

Roy, S. a. and Alloway, K. D. (2001). Coincidence detection or temporal integration? What the neurons in somatosensory cortex are doing. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(7):2462–73.

Ruiz, O. and Paradiso, M. a. (2012). Macaque V1 representations in natural and reduced visual contexts: spatial and temporal properties and influence of saccadic eye movements. *Journal of neurophysiology*, 108(1):324–33.

Rust, N. C., Mante, V., Simoncelli, E. P., and Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature neuroscience*, 9(11):1421–31.

Rust, N. C. and Movshon, J. A. (2005). In praise of artifice. *Nature neuroscience*, 8(12):1647–50.

Saleem, A. B., Ayaz, A., Jeffery, K. J., Harris, K. D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. *Nature neuroscience*, (November).

Sanada, T. M. and Ohzawa, I. (2006). Encoding of three-dimensional surface slant in cat visual areas 17 and 18. *Journal of neurophysiology*, 95(5):2768–86.

Sanderson, K. J. (1971). The projection of the visual field to the lateral geniculate and medial interlaminar nuclei in the cat. *The Journal of comparative neurology*, 143(1):101–8.

Saul, A. B. and Humphrey, A. L. (1990). Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *Journal of neurophysiology*, 64(1):206–24.

Schreiber, S., Fellous, J. M., Whitmer, D., Tiesinga, P., and Sejnowski, T. J. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54:925–931.

Schwartz, G., Taylor, S., Fisher, C., Harris, R., and Berry, M. J. (2007). Synchronized firing among retinal ganglion cells signals motion reversal. *Neuron*, 55(6):958–69.

Seelig, J. D. and Jayaraman, V. (2013). Feature detection and orientation tuning in the Drosophila central complex. *Nature*.

Seung, H. S. and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences of the United States of America*, 90(22):10749–53.

Shapley, R. and Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual review of neuroscience*, 8(1978):547–83.

Siddiqui, M. S. M. and Bhaumik, B. (2013). A study on surface slant encoding in V1. *Frontiers in systems neuroscience*, 7(November):87.

Simoncelli, E. P. and Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision research*, 38(5):743–61.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, 24:1193–216.

Smith, a. T., Wall, M. B., Williams, a. L., and Singh, K. D. (2006). Sensitivity to optic flow in human cortical areas MT and MST. *The European journal of neuroscience*, 23(2):561–9.

Somers, D. C., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 15(8):5448–65.

Sompolinsky, H. and Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. *Current opinion in neurobiology*, 7(4):514–22.

Stanley, G. B. (2013). Reading and writing the neural code. *Nature neuroscience*, 16(3):259–63.

Stanley, G. B., Jin, J., Wang, Y., Desbordes, G., Wang, Q., Black, M. J., and Alonso, J.-M. (2012). Visual orientation and directional selectivity through thalamic synchrony. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(26):9073–88.

Stoelzel, C. R., Bereshpolova, Y., and Swadlow, H. a. (2009). Stability of thalamocortical synaptic transmission across awake brain states. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(21):6851–9.

Tanaka, K. (1983). Cross-correlation analysis of geniculostriate neuronal relationships in cats. *Journal of neurophysiology*, 49(6):1303–18.

Tolhurst, D. J., Movshon, J. A., and Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental brain research.*, 41(3-4):414–9.

Tyler, C. W. and Sutter, E. E. (1979). Depth from spatial frequency difference: an old kind of stereopsis? *Vision research*, 19(8):859–65.

Usrey, W. M., Alonso, J. M., and Reid, R. C. (2000). Synaptic interactions between thalamic inputs to simple cells in cat visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(14):5461–7.

van Santen, J. P. and Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America. A, Optics and image science*, 2(2):300–21.

Vogels, R., Spileers, W., and Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental brain research.*, 77(2):432–6.

Wang, H.-P., Spencer, D., Fellous, J.-M., and Sejnowski, T. J. (2010a). Synchrony of thalamocortical inputs maximizes cortical reliability. *Science (New York, N.Y.)*, 328(5974):106–9.

Wang, Q., Webber, R. M., and Stanley, G. B. (2010b). Thalamic synchrony and the adaptive gating of information flow to cortex. *Nature neuroscience*, 13(12):1534–41.

Wehr, M. and Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442–6.

Weliky, M., Fiser, J., Hunt, R. H., and Wagner, D. N. (2003). Coding of natural scenes in primary visual cortex. *Neuron*, 37(4):703–18.

Weng, C., Yeh, C.-I., Stoelzel, C. R., and Alonso, J.-M. (2005). Receptive field size and response latency are correlated within the cat visual thalamus. *Journal of neurophysiology*, 93(6):3537–47.

Werblin, F. S. (2011). The retinal hypercircuit: a repeating synaptic interactive motif underlying visual function. *The Journal of physiology*, 589(Pt 15):3691–702.

Wilent, W. B. and Contreras, D. (2005). Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nature neuroscience*, 8(10):1364–70.

Wissig, S. C. and Kohn, A. (2012). The influence of surround suppression on adaptation effects in primary visual cortex. *Journal of neurophysiology*, pages 3370–3384.

Xie, X. (2002). Threshold behaviour of the maximum likelihood method in population decoding. *Network (Bristol, England)*, 13(4):447–56.

Yarrow, S., Challis, E., and Seriès, P. (2012). Fisher and Shannon information in finite neural populations. *Neural computation*, 24(7):1740–80.

Zhang, K. and Sejnowski, T. J. (1999). Neuronal Tuning: To Sharpen or Broaden? *Neural Computation*, 11(1):75–84.